PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

Sistema de Tomada de Decisão na Fase de Pré-Matrícula em uma Universidade Particular.

Isabela Braune Grego

PROJETO FINAL DE GRADUAÇÃO

CENTRO TÉCNICO E CIENTÍFICO – CTC
DEPARTAMENTO DE INFORMÁTICA
Curso de Graduação em Ciência da Computação

Rio de Janeiro, Dezembro de 2022

Isabela Braune Grego

Sistema de Tomada de Decisão na Fase de Pré-Matrícula em uma Universidade Particular.

Relatório de Projeto Final, apresentado ao programa Curso Ciência da Computação da PUC-Rio como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Marcos Vianna Villas

Rio de Janeiro, Dezembro de 2022

Agradecimentos

O desenvolvimento deste projeto contou com a ajuda de diversas pessoas, dentre as quais agradeço:

À minha família e à Nathalia Hinz por todo apoio e incentivo que eles sempre me deram.

Aos meus amigos Victor Martins, Giovanna Bottino e Taisa Felix que eu tive o prazer de conhecer na PUC-Rio e que sempre estiveram dispostos a me ajudar ao longo de curso.

Ao meu orientador Marcos Villas que me acompanhou pontualmente e me deu todo o auxílio necessário para a criação deste projeto.

Resumo

Grego, Isabela Braune; Villas, Marcos Vianna. Sistema de Tomada de Decisão na Fase de Pré-Matrícula em uma Universidade Particular. Rio de Janeiro, 2022. 44p. Relatório de Projeto Final – Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro.

Este projeto teve como objetivo solucionar o problema da falta de vagas no período de matrícula dos alunos.

Logo, foi necessário avaliar como é realizada a previsão da quantidade de vagas por disciplina e a alocação de professores hoje na PUC-Rio e procurar uma solução para estes problemas.

Este projeto visa apoiar os coordenadores dos cursos de informática na tomada de decisão de modo a oferecer mais turmas para mitigar o problema da falta de vagas, bem como a alocação dos professores.

Palavras-chave: Alocação de professores; Modelos preditivos; Quantidade de vagas; Power BI; Google Cloud Platform; Notebook Python; Previsão da quantidade de vagas por disciplina; Tomada de decisão; Pré-Matrícula;

Abstract

Grego, Isabela Braune; Villas, Marcos Vianna. Decision-making System in the Pre-Enrollment Phase at a Private University. Rio de Janeiro, 2022. 44p. Final Project Report – IT Department. Pontifícia Universidade Católica do Rio de Janeiro.

This project aimed to solve the problem of lack of vacancies during the enrollment period for students.

Therefore, it was necessary to evaluate how the prediction of the number of vacancies per subject and the allocation of professors today at PUC-Rio is carried out and to look for a solution to these problems.

This project aims to support the coordinators of computer courses in decision-making in order to offer more classes to mitigate the problem of lack of vacancies, as well as the allocation of professors.

Keywords: Allocation of teachers; Predictive models; Number of vacancies; Power BI; Google Cloud Platform; Python notebook; Forecasting the number of vacancies per subject; Decision making; Pre-Registration;

Sumário

Índice das figuras

Índice das tabelas

1. Introdução

O processo de matrícula em uma universidade é crucial para o aluno, pelo fato de que é chegado o momento em que é necessário organizar a grade de horários com as disciplinas que ele deseja cursar naquele período.

Ao longo da minha graduação, identifiquei não só a minha dificuldade como também a de outros alunos na fase de matrícula, em virtude da falta de vagas disponíveis em algumas disciplinas. Dificuldade esta que me incentivou a procurar uma solução para tal problema. Após conversar com o coordenador do Curso de Ciência da Computação da PUC-Rio, professor Ivan Mathias Filho, detectei dois desafios a serem vencidos pelas equipes de coordenação do Departamento de Informática na fase de matrícula. O primeiro desafio é definir a quantidade de vagas por disciplina a ser disponibilizada em cada período. O segundo desafio é remanejar os professores das turmas existentes, avaliando assim a possibilidade de criação de novas turmas e quem pode lecioná-las.

O Departamento de Informática da PUC-Rio é composto por duas coordenações. A coordenação das graduações de Ciência da Computação e de Sistemas de Informação, que é coordenada pelo professor Ivan Mathias Filho, e a coordenação da graduação em Engenharia da Computação, coordenada pelo professor Augusto Baffa.

Portanto, o projeto do Sistema de Tomada de Decisão na Fase de Pré-Matrícula em uma Universidade Particular visa apoiar as coordenações dos cursos de Informática da PUC-Rio nos dois desafios, de modo a oferecer mais turmas, mitigando o problema de falta de vagas, bem como o de alocação de professores.

2. Situação Atual

Após detectar dois dos desafios da equipe de coordenação do Curso de Ciência da Computação da PUC-Rio na fase de matrícula, que consiste em calcular a quantidade de vagas por disciplina disponibilizada e remanejar os professores das turmas avaliando a possibilidade de criação de novas turmas em cada período, foi necessário procurar entender como esses desafios são enfrentados hoje.

Ao consultar a secretária do Departamento de Informática da PUC-Rio, Fernanda Basílio, sobre como é feito o planejamento da quantidade de vagas disponibilizadas em cada período, constatei que ele é feito manualmente pela coordenação do departamento através de relatórios, apresentados na Seção 5 - Levantamento e Entendimento dos Dados do Projeto, emitidos pelo Decanato do Centro Técnico Científico da PUC-Rio. O Decanato faz toda a estruturação dos cursos que compõem o CTC e administra um sistema que possui as turmas oferecidas por curso, e que, ao final, gera estes relatórios com dados históricos dos últimos seis períodos, que são disponibilizados para a equipe de coordenação realizar o planejamento.

Além disso, em uma consulta ao professor Ivan Mathias Filho, foi possível inferir que o remanejamento dos professores das turmas ainda não é realizado, pois não existe um controle adequado dos horários disponíveis de cada professor e quais disciplinas ele pode lecionar. O que a coordenação do curso consegue fazer é avaliar se um professor tem disponibilidade para lecionar uma turma, sem verificar todas as possibilidades dentre todos os professores do Departamento de Informática.

3. Proposta e Objetivos do trabalho

Esse trabalho teve como objetivo auxiliar os coordenadores do Departamento de Informática no desafio da previsão de vagas e de turmas, bem como a alocação de professores, visando sinalizar aos coordenadores a demanda dos alunos na fase de pré-matrícula nas disciplinas do Departamento de Informática.

Desta forma, este projeto final consiste no uso de dados fornecidos pela secretaria do Departamento de Informática e na simulação dos dados de matrícula, disciplinas e professores feita por mim. Esses dados são utilizados para o desenvolvimento de uma solução de Business Intelligence que coleta e trata esses dados em uma ferramenta de BI. A partir desse tratamento, esses dados são utilizados no desenvolvimento do modelo preditivo e de alocação de recurso, especificados nas seções 7 - Ferramentas de Business Intelligence e 8 - Modelos de Machine Learning, e carregados em uma ferramenta de visualização para a apresentação dessas informações de forma clara para os usuários com o propósito de mitigar os desafios da coordenação.

Após o tratamento dos dados e a criação dos modelos, o coordenador poderá acessar a ferramenta de visualização com as informações necessárias para que ele seja capaz de tomar decisões em relação à quantidade de turmas que serão disponibilizadas e a alocação dos professores nessas turmas.

É possível conferir a arquitetura desse projeto na seção 5 - Arquitetura da Solução, onde estão especificadas as fontes de dados, o fluxo da informação e o resultado que será fornecido aos usuários.

É importante salientar a importância do armazenamento dos dados da fase de pré-matrícula. Este projeto foi desenvolvido com os dados de matrícula e simulação de dados, pois foi aferido que as informações coletadas na fase de pré-matrícula, também chamada de simulação da matrícula, da PUC-Rio não são armazenadas. Com esses dados, seria possível ter um histórico da quantidade de alunos que solicitaram vagas em cada disciplina e desenvolver o modelo em cima desses dados, com o intuito de mitigar a falta de vagas na universidade.

Este projeto tem como um de seus objetivos prever a quantidade de vagas das discplinas, o que promove o aumento da quantidade de turmas e afeta a disponibilidade das salas da universidade. A disponibilidade das salas gera uma limitação à criação de turmas e considerar esta informação está fora do escopo deste projeto.

4. Metodologia

O plano de ação foi constituído de: (1) levantamento e entendimento dos dados que são utilizados pela equipe de coordenação do Departamento de Informática da PUC-Rio na previsão da quantidade de vagas que serão disponibilizadas por período; (2) mapeamento dos dados do projeto; (3) estudo das ferramentas de Business Intelligence[1] e dos modelos de Machine Learning[2]; (4) estudo das ferramentas de visualização para o desenvolvimento do dashboard para os coordenadores dos cursos do DI; (5) definição do critério de qualidade de teste dos modelos; (6) definição da arquitetura da solução; (7) desenvolvimento do ETL; (8) desenvolvimento dos modelos; (9) teste dos modelos; (10) definição do modelo multidimensional utilizado na ferramenta de visualização; (11) criação das visualizações com os dados finais dos modelos; (12) apresentação do dashboard para os usuários finais (coordenadores do Departamento de Informática); (13) consolidação dos comentários e impressões dos coordenadores.

[1] Business Intelligence é um conjunto de métodos de extração, tratamento, análise e visualização de dados para apoiar nas decisões de negócio.
[2] Machine Learning é uma tecnologia capaz de aprender padrões através de treinamentos e testes de um grande volume de dados.

5. Arquitetura da Solução

A arquitetura da solução deste projeto (Figura 5.1) considera a existência de seis etapas, desde a captura dos dados até a visualização no dashboard.

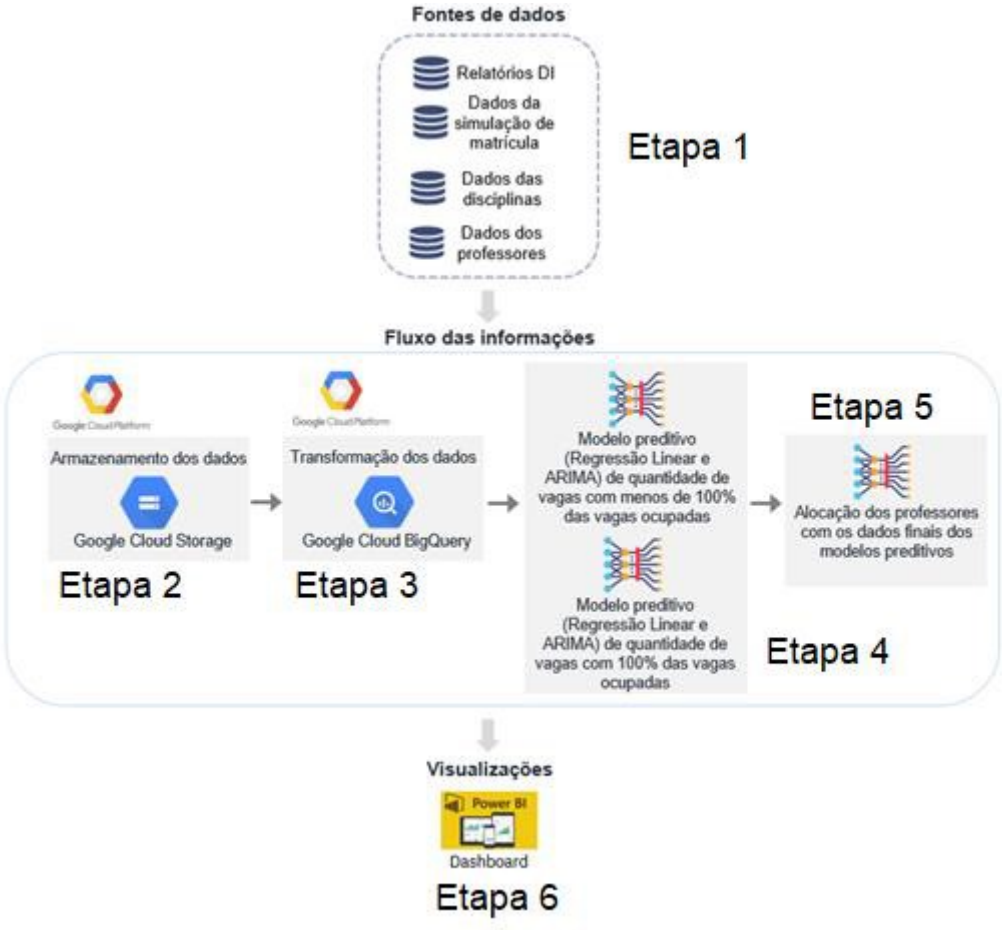


Figura 5.1: Arquitetura da solução.

A primeira etapa consiste na captura e simulação dos dados de matrícula, professores e disciplinas. Essa etapa está detalhada na seção 6 - Levantamento e entendimento dos dados do projeto.

A segunda e a terceira etapa compõem o processo de carga e transformação dos dados. Na segunda etapa ocorre o armazenamento dos dados no Google Cloud Storage (GCS) e a terceira etapa é o processo de tranformação desses dados no Google Cloud Bigquery (GCB). Essas duas etapas estão especificadas na seção 8 - Ferramentas de Business Intelligence.

A quarta etapa desta arquitetura consiste nos modelos preditivos de quantidade de vagas com menos de 100% das vagas ocupadas e nos modelos preditivos de quantidade de vagas com 100% das vagas ocupadas. Essa etapa está detalhada na seção 9 - Modelos de Machine Learning.

A alocação de professores é a quinta etapa da arquitetura e ela foi desenvolvida com os dados finais dos modelos preditivos. O detalhamento deste desenvolvimento está na seção 10 - Alocação de recursos.

A última etapa é o dashboard que os coordenadores do Departamento de Informática poderão consultar para a tomada de decisões. Essa etapa está especificada na seção 11 - Visualização.

6. Levantamento e entendimento dos dados do projeto

O primeiro passo do projeto do Sistema de Tomada de Decisão na Fase de Pré-Matrícula em uma Universidade Particular foi entrar em contato com a equipe da Coordenação do Departamento de Informática da PUC-Rio e entender como hoje é calculada a quantidade de vagas disponibilizadas por disciplina em cada período.

Foram identificadas 5 tabelas/relatórios:

- 20212-Cursos do Dept. de Informática-CTC – dados reais das disciplinas que foram oferecidas em 2021.
- QueroCursarCTC_20212 – dados reais das disciplinas que os alunos tentaram se matricular na primeira fase de matrícula, mas não conseguiram.
- Vagas_Disciplina – dados reais com o histórico de vagas oferecidas e vagas ocupadas dos últimos cinco anos.
- Horários disponíveis dos professores – dados simulados dos horários que os professores têm disponível.
- Disciplinas que os professores lecionam – dados simulados indicando as disciplinas que os professores lecionam.

Cada uma destas tabelas/relatórios é descrita a seguir.

A partir disso, entrei em contato com a secretária do Departamento de Informática, Fernanda Basílio, que encaminhou dois relatórios para mim. O primeiro relatório se chama 20212-Cursos do Dept. de Informática-CTC (Figura 6.1 e 6.2). Ele contém as informações das disciplinas que foram oferecidas no último semestre par, ou seja, no segundo período de 2021, com a quantidade de vagas, bloqueios e a série histórica de ocupação de vagas (Tabela 6.1).

Campo	Definição
Modalidade	Indica se é uma disciplina de modalidade online ou presencial.
Código	Indica o código da disciplina.
Nome	Indica o nome da disciplina.
Turma	Indica a turma da disciplina.
Dia	Indica o dia da semana em que a disciplina é lecionada.
Início	Indica que horas a aula começa.
Fim	Indica que horas a aula termina.
Sala	Indica em qual sala ficará a turma.
Tipo Sala	Indica o tipo da sala em que a aula será dada (Teoria, Laboratório).
Modo Horário	Indica se será uma disciplina com aula online ou presencial.
Professor	Indica o nome do professor que leciona a disciplina.
Horista	Indica se o professor é do quadro titular ou não.

Comp	Indica se a turma teve junção de turmas de códigos diferentes, ou até de códigos iguais, com o mesmo professor em um mesmo horário.
Eletiva	Indica se a disciplina é eletiva.
Total	Quantidade total de vagas oferecidas naquela turma.
CAL	Quantidade de vagas disponíveis para os calouros.
CEG	Quantidade de vagas disponíveis para qualquer curso de engenharia.
COP	Quantidade de vagas disponíveis exclusivamente para o curso de engenharia da computação.
INF	Quantidade de vagas disponíveis exclusivamente para o curso de informática.
QQC	Quantidade de vagas disponíveis para alunos de qualquer curso da PUC.
20192	Quantidade de vagas ocupadas pelos alunos pela quantidade de vagas totais da disciplina no segundo período de 2019.
20201	Quantidade de vagas ocupadas pelos alunos pela quantidade de vagas totais da disciplina no primeiro período de 2020.
20202	Quantidade de vagas ocupadas pelos alunos pela quantidade de vagas totais da disciplina no segundo período de 2020.
20211	Quantidade de vagas ocupadas pelos alunos pela quantidade de vagas totais da disciplina no primeiro período de 2021.
20212	Quantidade de vagas ocupadas pelos alunos pela quantidade de vagas totais da disciplina no segundo período de 2021.
20221	Quantidade de vagas ocupadas pelos alunos pela quantidade de vagas totais da disciplina no primeiro período de 2022.
Média Ocupação	Quantidade média de alunos que se matricularam na disciplina nos seis meses anteriores ao cálculo da previsão de matrícula.
Média Cancelamento	Quantidade média de alunos que cancelaram a disciplina até a fase de cancelamento (um pouco depois da G1) nos seis meses anteriores ao cálculo da previsão de matrícula.

Tabela 6.1: Descrição de dados do relatório 20212-Cursos do Dept. de Informática-CTC realizado por mim.

Modalidade	Código	Nome	Turma	Dia	Início	Fim	Sala	Tipo Sala	Modo Horário	Professor	Horista	Comp	Eletiva
ONLINE	ADM1006	. A ANAL. ADMINISTR	2WA	3	15	17	L504	TEORIA	ONLINE	ARDO NOLLA R	S	N	N
ONLINE	ADM1006	. A ANAL. ADMINISTR	2WA	5	15	17	L504	TEORIA	ONLINE	ARDO NOLLA R	S	N	N
ONLINE	ADM1019	INTR A FINANÇAS	2WA	3	13	15	L508	TEORIA	ONLINE	ARDO NOLLA R	S	N	N
ONLINE	ADM1019	INTR A FINANÇAS	2WA	5	13	15	L504	TEORIA	ONLINE	ARDO NOLLA R	S	N	N
ONLINE	ENG1132	UAÇÃO ENGENHARIA	3VA	0	0	0				LA ROCQUE R	N	N	N
ONLINE	ENG1133	UAÇÃO ENGENHARIA I	3VA	0	0	0				LA ROCQUE R	N	N	N
ONLINE	ENG1153	IONADO ENGENHARI	3VA	0	0	0				SA LEMOS CAV	S	N	N
ONLINE	INF1009	LÓGICA P/A COMP	3WA	3	13	15	L414	TEORIA	ONLINE	IS ENGLANDE	S	N	N
ONLINE	INF1009	LÓGICA P/A COMP	3WA	5	13	15	L144	TEORIA	ONLINE	IS ENGLANDE	S	N	N
ONLINE	INF1009	LÓGICA P/A COMP	3WB	3	15	17	L456	TEORIA	ONLINE	IS ENGLANDE	S	N	N
ONLINE	INF1009	LÓGICA P/A COMP	3WB	5	15	17	L456	TEORIA	ONLINE	IS ENGLANDE	S	N	N
ONLINE	INF1010	R DE DADOS AVANÇA	3WA	2	15	17	L546	AB. MICR	ONLINE	CESAR ESPIND	S	N	N
ONLINE	INF1010	R DE DADOS AVANÇA	3WA	4	15	17	L546	AB. MICR	ONLINE	CESAR ESPIND	S	N	N
ONLINE	INF1010	R DE DADOS AVANÇA	3WB	2	13	15		AB. MICR	ONLINE	RNANDO BESS	S	N	N
ONLINE	INF1010	R DE DADOS AVANÇA	3WB	4	13	15		AB. MICR	ONLINE	RNANDO BESS	S	N	N
ONLINE	INF1012	ODELAGEM DE DADO	3WA	6	13	15	L242	TEORIA	ONLINE	N MATHIAS FI	N	N	N

Figura 6.1: Primeira parte do relatório 20212-Cursos do Dept. de Informática-CTC.xlsx da coordenação do Departamento de Informática da PUC-Rio.

Total	CAL	CEG	COP	INF	QQC	20192	20201	20202	20211	20212	20221	Média Ocupação	Média Cancelamentos	Observações
30					30	4/30	0/0	7/30	0/0	13/30	0/0	8	0	
30					30	4/30	0/0	7/30	0/0	13/30	0/0	8	0	
30					30	27/30	25/30	34/35	27/30	26/30	24/30	27,17	2	
30					30	27/30	25/30	34/35	27/30	26/30	24/30	27,17	2	
30		30				13/30	9/30	13/30	17/30	12/30	8/8	12	1	
30		30				24/30	18/30	13/30	18/30	21/30	22/30	19,33	2	
30					30	12/30	13/30	11/30	12/30	28/30	17/30	15,5	1	
41	38			3		21/30	32/43	15/20	44/43	24/43	41/40	29,5	2	
41	38			3		21/30	32/43	15/20	44/43	24/43	41/40	29,5	2	
40					40	26/46	24/30	27/38	19/30	32/40	41/53	28,17	5	
40					40	26/46	24/30	27/38	19/30	32/40	41/53	28,17	5	
30					30	23/26	21/30	19/32	18/30	24/30	26/30	21,83	4	
30					30	23/26	21/30	19/32	18/30	24/30	26/30	21,83	4	
30					30	15/25	15/30	25/31	24/30	17/30	23/30	19,83	2	
30					30	15/25	15/30	25/31	24/30	17/30	23/30	19,83	2	
47	37				10	25/31	42/58	23/27	48/50	34/49	26/30	33	2	

Figura 6.2: Segunda parte do relatório 20212-Cursos do Dept. de Informática-CTC.xlsx da coordenação do Departamento de Informática da PUC-Rio.

O segundo relatório, QueroCursarCTC_20212 (Figura 6.3), contém as informações das disciplinas que os alunos tentaram se matricular na primeira fase de matrícula, mas

não conseguiram por falta de vaga. Ele possui campo indicando a disciplina e quantidade de vagas totais solicitadas pelos alunos (Tabela 6.2). A partir dessa planilha, a coordenação é capaz de ajustar a quantidade de vagas que serão disponibilizadas na segunda fase da matrícula.

Campo	Definição
Disciplina	Indica o código da disciplina solicitada pelo aluno.
curso	Indica o curso da disciplina.
habil.	Indica qual é o curso do aluno que solicitou a disciplina.
enfase	Indica se esse aluno está cursando o Ciclo Básico.
pedidos	Soma da quantidade de vagas que os alunos solicitaram para aquela disciplina.
MT_CB	Quantidade de alunos do Ciclo Básico que se matricularam nessa disciplina na primeira fase de matrícula.
MT_PROF	Quantidade de alunos do Ciclo Profissional que se matricularam nessa disciplina na primeira fase de matrícula
MT_OUT_CURSOS	Quantidade de alunos de Outros Cursos que se matricularam nessa disciplina na primeira fase de matrícula.

Tabela 6.2: Descrição de dados do relatório QueroCursarCTC_20212.xlsx realizado por mim.

Disciplina	curso	habil.	enfase	pedidos	MT_CB	MT_PROF	MT_OUT_CURSOS
ENG1003	CEG	BAB	CCB	5	40	0	0
ENG1003	CEG	BCL	CCB	2	40	0	0
ENG1003	CEG	BCO	CCB	11	40	0	0
ENG1003	CEG	BEL	CCB	1	40	0	0
ENG1003	CEG	BMN		1	40	0	0
ENG1003	CEG	BMN	CCB	1	40	0	0
ENG1003	CEG	BPD		2	40	0	0
ENG1003	CEG	BPD	CCB	37	40	0	0
ENG1003	CEG	BQM	CCB	6	40	0	0

Figura 6.3: Relatório QueroCursarCTC_20212.xlsx da coordenação do Departamento de Informática da PUC-Rio.

Tanto o relatório 20212-Cursos do Dept. de Informática-CTC quanto o QueroCursarCTC_20212 possuem 20212 em seus nomes, pois eles foram gerados no segundo período de 2021 e são utilizados para apoiar a previsão da quantidade de vagas que serão disponibilizadas pelos coordenadores no primeiro período de 2022.

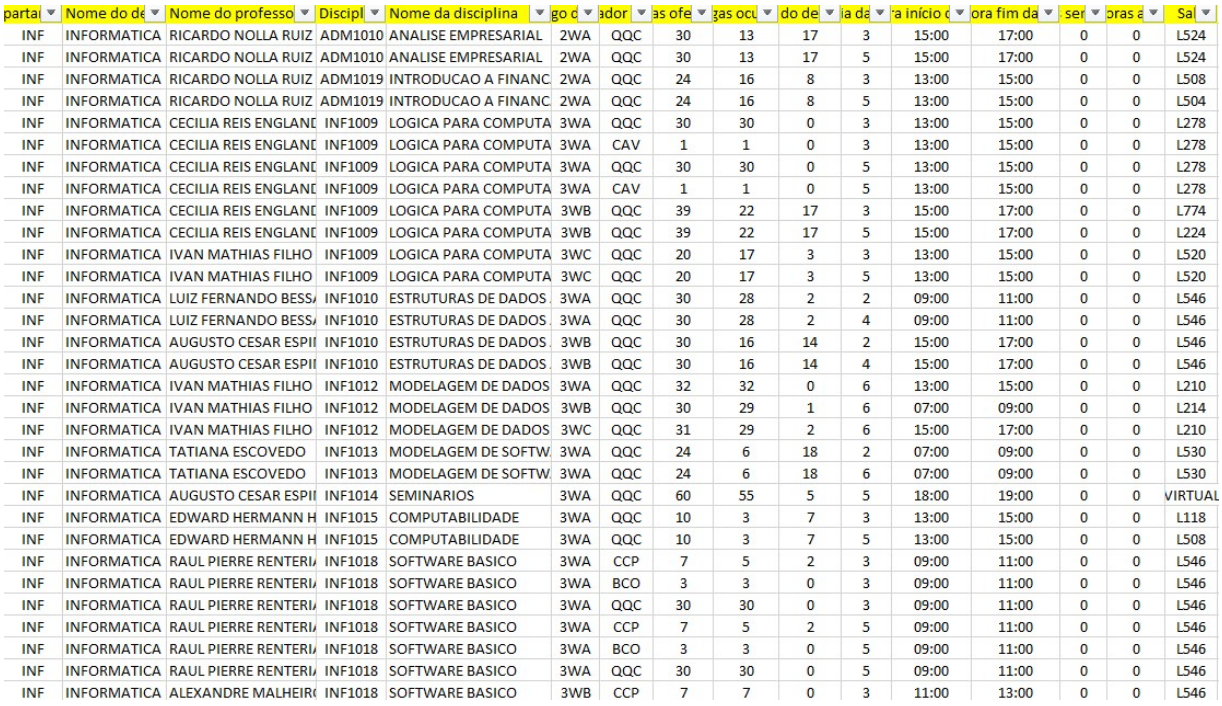
Após analisar esses dois relatórios, foi levantada a importância de utilizar dados reais complementados por dados simulados para a previsão de quantidade de vagas. Os dados reais são importantes para as disciplinas que não tiveram todas as vagas que foram oferecidas ocupadas, pois não há o risco de mais alunos terem o interesse de cursá-las e não terem conseguido pois foi atingida a quantidade máxima de vagas oferecidas. Os dados simulados são importantes para as disciplinas que tiveram a quantidade de vagas

ocupadas igual à quantidade de vagas oferecidas, pois é possível simular que a quantidade de alunos interessados na disciplina foi maior do que a quantidade de vagas oferecidas, assim, podendo testar no modelo esse caso.

Desta forma, solicitei o histórico de vagas das disciplinas e a secretaria do Departamento de Informática disponibilizou o relatório Vagas_Disciplina (Figura 6.4), que é um relatório gerado no sistema SAU intranet voltado para os funcionários administrativos, que tem o histórico de vagas oferecidas e vagas ocupadas na PUC-Rio de 2017 a 2022 (Tabela 6.3).

Campo	Definição
Departamento	Indica o código do departamento
Nome do departamento	Indica o nome do departamento
Nome do professor	Indica o nome do professor
Disciplina	Indica o código da disciplina
Nome da disciplina	Indica o nome da disciplina
Código da turma	Indica o código da turma
Identificador de bloqueio	Indica se a disciplina tem algum bloqueio
Vagas oferecidas	Indica a quantidade de vagas oferecidas
Vagas ocupadas	Indica a quantidade de vagas ocupadas
Saldo de vagas	Indica o saldo de vagas
Dia da aula	Indica o dia da aula
Hora início da aula	Indica o horário de início da aula
Hora fim da aula	Indica o horário de término da aula
Qtd. horas sem horário fixo	Indica a quantidade de horas sem horário fixo
Qtd. horas a distância	Indica a quantidade de horas à distância
Sala	Indica a sala

Tabela 6.3: Descrição de dados do relatório Vagas_Disciplina.xlsx realizado por mim.



parta	Nome do de	Nome do professo	Discipl	Nome da disciplina	go d	ador	as ofe	gas ocu	do de	ia da	ra início d	ora fim da	ser	oras a	Sa
INF	INFORMATICA	RICARDO NOLLA RUIZ	ADM1010	ANALISE EMPRESARIAL	2WA	QQC	30	13	17	3	15:00	17:00	0	0	L524
INF	INFORMATICA	RICARDO NOLLA RUIZ	ADM1010	ANALISE EMPRESARIAL	2WA	QQC	30	13	17	5	15:00	17:00	0	0	L524
INF	INFORMATICA	RICARDO NOLLA RUIZ	ADM1019	INTRODUCAO A FINANC	2WA	QQC	24	16	8	3	13:00	15:00	0	0	L508
INF	INFORMATICA	RICARDO NOLLA RUIZ	ADM1019	INTRODUCAO A FINANC	2WA	QQC	24	16	8	5	13:00	15:00	0	0	L504
INF	INFORMATICA	CECILIA REIS ENGLAND	INF1009	LOGICA PARA COMPUTA	3WA	QQC	30	30	0	3	13:00	15:00	0	0	L278
INF	INFORMATICA	CECILIA REIS ENGLAND	INF1009	LOGICA PARA COMPUTA	3WA	CAV	1	1	0	3	13:00	15:00	0	0	L278
INF	INFORMATICA	CECILIA REIS ENGLAND	INF1009	LOGICA PARA COMPUTA	3WA	QQC	30	30	0	5	13:00	15:00	0	0	L278
INF	INFORMATICA	CECILIA REIS ENGLAND	INF1009	LOGICA PARA COMPUTA	3WA	CAV	1	1	0	5	13:00	15:00	0	0	L278
INF	INFORMATICA	CECILIA REIS ENGLAND	INF1009	LOGICA PARA COMPUTA	3WB	QQC	39	22	17	3	15:00	17:00	0	0	L774
INF	INFORMATICA	CECILIA REIS ENGLAND	INF1009	LOGICA PARA COMPUTA	3WB	QQC	39	22	17	5	15:00	17:00	0	0	L224
INF	INFORMATICA	IVAN MATHIAS FILHO	INF1009	LOGICA PARA COMPUTA	3WC	QQC	20	17	3	3	13:00	15:00	0	0	L520
INF	INFORMATICA	IVAN MATHIAS FILHO	INF1009	LOGICA PARA COMPUTA	3WC	QQC	20	17	3	5	13:00	15:00	0	0	L520
INF	INFORMATICA	LUIZ FERNANDO BESSA	INF1010	ESTRUTURAS DE DADOS	3WA	QQC	30	28	2	2	09:00	11:00	0	0	L546
INF	INFORMATICA	LUIZ FERNANDO BESSA	INF1010	ESTRUTURAS DE DADOS	3WA	QQC	30	28	2	4	09:00	11:00	0	0	L546
INF	INFORMATICA	AUGUSTO CESAR ESPII	INF1010	ESTRUTURAS DE DADOS	3WB	QQC	30	16	14	2	15:00	17:00	0	0	L546
INF	INFORMATICA	AUGUSTO CESAR ESPII	INF1010	ESTRUTURAS DE DADOS	3WB	QQC	30	16	14	4	15:00	17:00	0	0	L546
INF	INFORMATICA	IVAN MATHIAS FILHO	INF1012	MODELAGEM DE DADOS	3WA	QQC	32	32	0	6	13:00	15:00	0	0	L210
INF	INFORMATICA	IVAN MATHIAS FILHO	INF1012	MODELAGEM DE DADOS	3WB	QQC	30	29	1	6	07:00	09:00	0	0	L214
INF	INFORMATICA	IVAN MATHIAS FILHO	INF1012	MODELAGEM DE DADOS	3WC	QQC	31	29	2	6	15:00	17:00	0	0	L210
INF	INFORMATICA	TATIANA ESCOVEDO	INF1013	MODELAGEM DE SOFTW	3WA	QQC	24	6	18	2	07:00	09:00	0	0	L530
INF	INFORMATICA	TATIANA ESCOVEDO	INF1013	MODELAGEM DE SOFTW	3WA	QQC	24	6	18	6	07:00	09:00	0	0	L530
INF	INFORMATICA	AUGUSTO CESAR ESPII	INF1014	SEMINARIOS	3WA	QQC	60	55	5	5	18:00	19:00	0	0	VIRTUAL
INF	INFORMATICA	EDWARD HERMANN H	INF1015	COMPUTABILIDADE	3WA	QQC	10	3	7	3	13:00	15:00	0	0	L118
INF	INFORMATICA	EDWARD HERMANN H	INF1015	COMPUTABILIDADE	3WA	QQC	10	3	7	5	13:00	15:00	0	0	L508
INF	INFORMATICA	RAUL PIERRE RENTERIA	INF1018	SOFTWARE BASICO	3WA	CCP	7	5	2	3	09:00	11:00	0	0	L546
INF	INFORMATICA	RAUL PIERRE RENTERIA	INF1018	SOFTWARE BASICO	3WA	BCO	3	3	0	3	09:00	11:00	0	0	L546
INF	INFORMATICA	RAUL PIERRE RENTERIA	INF1018	SOFTWARE BASICO	3WA	QQC	30	30	0	3	09:00	11:00	0	0	L546
INF	INFORMATICA	RAUL PIERRE RENTERIA	INF1018	SOFTWARE BASICO	3WA	CCP	7	5	2	5	09:00	11:00	0	0	L546
INF	INFORMATICA	RAUL PIERRE RENTERIA	INF1018	SOFTWARE BASICO	3WA	BCO	3	3	0	5	09:00	11:00	0	0	L546
INF	INFORMATICA	RAUL PIERRE RENTERIA	INF1018	SOFTWARE BASICO	3WA	QQC	30	30	0	5	09:00	11:00	0	0	L546
INF	INFORMATICA	ALEXANDRE MALHEIRO	INF1018	SOFTWARE BASICO	3WB	CCP	7	7	0	3	11:00	13:00	0	0	L546

Figura 6.4: Relatório Vagas_Disciplina.xlsx da coordenação do Departamento de Informática da PUC-Rio.

Além desses dados encaminhados pelo DI, realizei a simulação dos dados de vaga das disciplinas para o caso de serem solicitadas um número maior de vagas na fase de pré-matrícula do que na liberação da quantidade de vagas oferecidas.

Para a simulação desses dados, eu reaproveitei a planilha Vagas_Disciplina.xlsx, renomeei a coluna “Vagas ocupadas” para “Vagas solicitadas” e utilizei a função “RANDBETWEEN” do excel para inserir números aleatórios de 0 a 50 nessa coluna. Eu utilizei o intervalo de 0 a 50, pois o número mínimo de vagas ocupadas da planilha Vagas_Disciplina é 2 e o número máximo é 45, então escolhi explorar esses limites com esse intervalo.

Além disso, realizei a simulação dos dados dos horários disponíveis dos professores (Figura 6.5 e Tabela 6.4) e dos dados das disciplinas que os professores lecionam (Figura 6.6 e Tabela 6.5) para utilizá-los para o desenvolvimento do modelo de alocação dos professores.

Campo	Definição
nome_professor	Indica o nome do professor
horario	Indica o horario que o professor tem disponível para dar aula

Tabela 6.4: Descrição de dados dos horários disponíveis dos professores realizado por mim.

nome_professor	horario
LUIZ FERNANDO BESSA SEIBEL	9-11 2a 4a
AUGUSTO CESAR ESPINDOLA BAFFA	15-17 2a 4a
EDWARD HERMANN HAEUSLER	13-15 3a 5a
EDWARD HERMANN HAEUSLER	11-13 2a 4a
ALESSANDRO FABRICIO GARCIA	15-17 3a 5a
ANDERSON OLIVEIRA DA SILVA	9-11 3a 5a
HELIO CORTES VIEIRA LOPES	7-9 2a 4a
LUIZ FERNANDO BESSA SEIBEL	17-19 3a 5a
MARCOS VIANNA VILLAS	7-9 3a 5a
SERGIO LIFSCHITZ	15-17 3a 5a
SIMONE DINIZ JUNQUEIRA BARBOSA	11-13 3a 5a
SIMONE DINIZ JUNQUEIRA BARBOSA	13-15 3a 5a
ALEXANDRE MALHEIROS MESLIN	17-19 2a 4a
IVAN MATHIAS FILHO	7-9 3a 5a
IVAN MATHIAS FILHO	7-9 2a 4a
WALDEMAR CELES FILHO	9-11 3a 5a

Figura 6.5: Relatório dos dados dos horários disponíveis dos professores simulados por mim.

Campo	Definição
nome_professor	Indica o nome do professor
nome_disciplina	Indica o nome da disciplina que o professor leciona

Tabela 6.5: Descrição de dados das disciplinas que os professores lecionam realizado por mim.

nome_professor	nome_disciplina
LUIZ FERNANDO BESSA SEIBEL	ESTRUTURAS DE DADOS AVANCADAS
AUGUSTO CESAR ESPINDOLA BAFFA	ESTRUTURAS DE DADOS AVANCADAS
EDWARD HERMANN HAEUSLER	COMPUTABILIDADE
EDWARD HERMANN HAEUSLER	ANALIS LEXICOS E SINTATICOS
ALESSANDRO FABRICIO GARCIA	PROJETO E CONSTRUCAO SISTEMAS
ANDERSON OLIVEIRA DA SILVA	INT ARQUITETURA COMPUTADORES
HELIO CORTES VIEIRA LOPES	INTRODUCAO A CIENCIA DOS DADOS
LUIZ FERNANDO BESSA SEIBEL	BANCO DE DADOS II
MARCOS VIANNA VILLAS	BANCO DE DADOS II
SERGIO LIFSCHITZ	BANCOS DE DADOS
SIMONE DINIZ JUNQUEIRA BARBOSA	INTR INT HUMANO-COMPUTADOR
ALEXANDRE MALHEIROS MESLIN	PROGRAMACAO PARA A WEB
IVAN MATHIAS FILHO	PROGRAMACAO ORIENTADA OBJETOS
WALDEMAR CELES FILHO	COMPUTACAO GRAFICA
AUGUSTO CESAR ESPINDOLA BAFFA	INTELIGENCIA ARTIFICIAL
MELISSA LEMOS CAVALIERE	ESTAGIO SUPERVISIONADO

Figura 6.6: Relatório dos dados das disciplinas que os professores lecionam simulados por mim.

7. Mapeamento dos dados do projeto

Após o recebimento dos relatórios e o entendimento dos dados que são atualmente utilizados para a previsão da quantidade de vagas que serão disponibilizadas para cada disciplina, foi feito o levantamento de todas as informações que serão necessárias para o desenvolvimento do modelo preditivo da quantidade de vagas e do modelo de alocação dos professores. Com isso, foram geradas as descrições dos dados (Tabela 7.1, Tabela 7.2 e Tabela 7.3) que será utilizado no desenvolvimento do Projeto Final II e, com base nesse dicionário, foram criados os modelos ER (Figura 7.1) e Relacional (Figura 7.2).

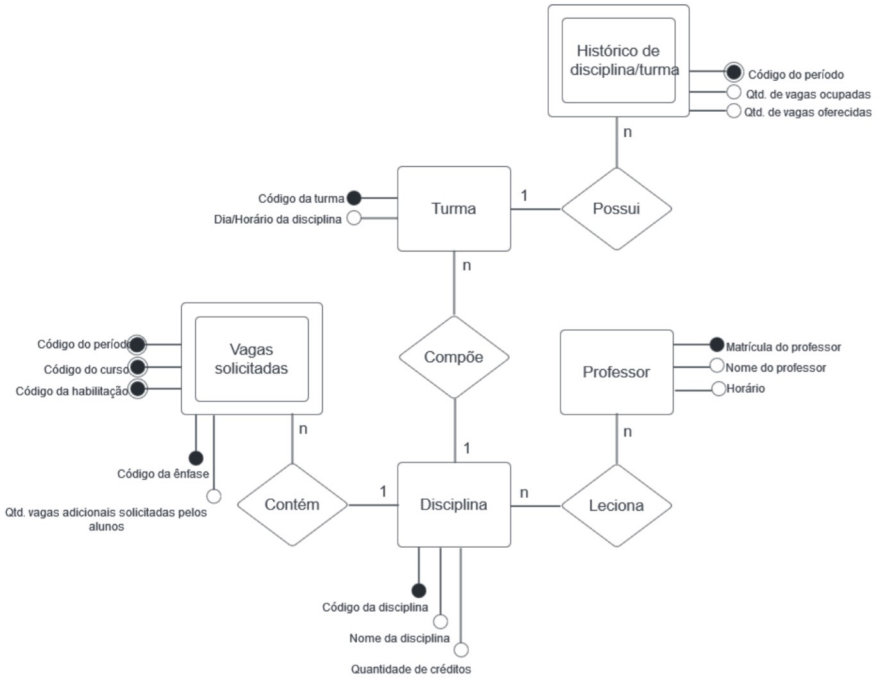


Figura 7.1: Modelo ER[3].

[3] Os atributos identificadores das entidades fracas e das entidades fortes estão representados no Modelo ER com uma bolota preenchida com a cor preta.

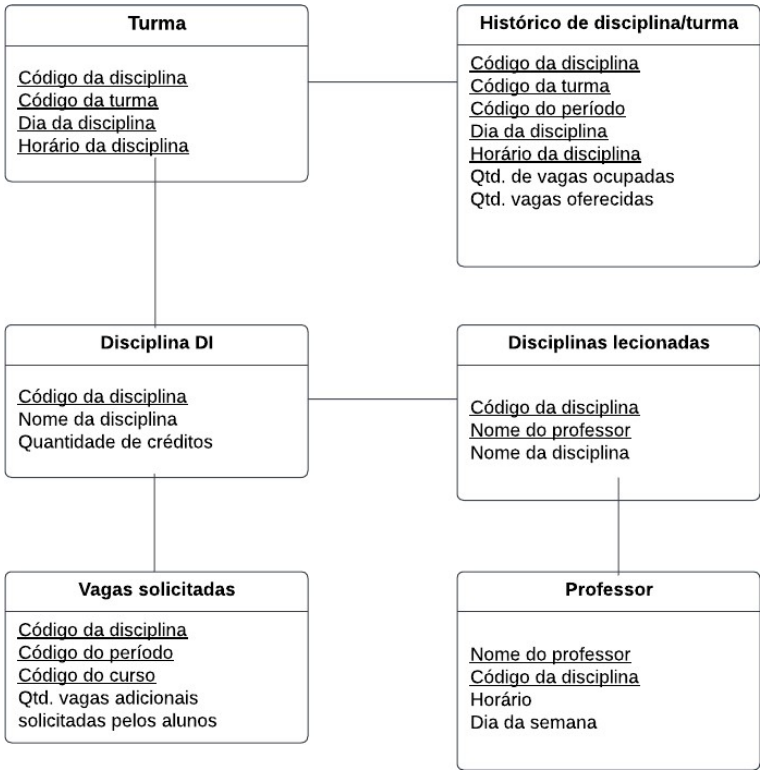


Figura 7.2: Modelo Relacional[4].

Campo	Tipo do dado	Definição
Código da disciplina	String	Indica o código da disciplina
Código da turma	String	Indica o código da turma
Código do período	String	Indica o código do período
Dia da disciplina	String	Indica o dia da disciplina
Horário da disciplina	String	Indica o horário da disciplina
Qtd. de vagas ocupadas	Inteiro	Indica a quantidade de vagas ocupadas
Qtd. de vagas oferecidas	Inteiro	Indica a quantidade de vagas oferecidas

Tabela 7.1: Descrição dos dados do histórico de disciplina/turma.

Campo	Tipo do dado	Definição
Código da disciplina	String	Indica o código da disciplina
Nome do professor	String	Indica o nome do professor
Nome da disciplina	String	Indica o nome da disciplina

[4] Os atributos identificadores estão representados no Modelo Relacional com sublinhado.

Quantidade de créditos	Inteiro	Indica a quantida de créditos

Tabela 7.2: Descrição dos dados históricos das disciplinas lecionadas.

Campo	Tipo do dado	Definição
Nome do professor	String	Indica o nome do professor
Código das disciplinas	String	Indica o código das disciplinas que o professor pode lecionar
Dias disponíveis do professor	String	Indica os dias que o professor pode dar aula
Horários disponíveis do professor		Indica os horários que o professor pode dar aula

Tabela 7.3: Descrição dos dados dos professores.

8. Ferramentas de Business Intelligence

O BI compreende vários métodos e tecnologias que permitem extrair, analisar, transformar e gerar visualizações para tomadas de decisão na área de negócio.

Para a criação de um sistema de Business Intelligence é necessário considerar o processo de captura, tratamento e carga dos dados (ETL – Extract, Transform, Load). Existem diversas ferramentas no mercado que são capazes de realizar esse processo, como o Oracle Data Integrator (ODI) e o Informatica Power Center. Porém, o projeto do Sistema de Tomada de Decisão na Fase de Pré-Matrícula, além de ser uma solução de BI, possui um modelo preditivo e um modelo de alocação de recursos. Desta forma, é indispensável avaliar aplicações que viabilizam a integração do processo de ETL com o Machine Learning e, portanto, essas ferramentas não são uma opção pois elas não solucionam todos os objetivos deste projeto.

Após alguns estudos, foi possível inferir que os serviços de plataformas na nuvem são a melhor solução para o desenvolvimento deste projeto, pois a partir de uma plataforma nuvem é possível integrar os serviços de Business Intelligence entre si e com modelos de Machine Learning.

Os serviços em nuvem [8] são provedores de plataformas (PaaS), softwares (SaaS) e infraestrutura (IaaS) que são acessados online. Eles trouxeram inúmeras vantagens para os desenvolvedores, pois cada serviço possui um valor específico, então você paga apenas pelos serviços que você está utilizando. Além disso, eles são de mais fácil acesso e não precisam de data centers físicos, o que diminui o custo de cada serviço.

A tecnologia está em constante evolução e hoje existem três principais plataformas de nuvem para desenvolvimento de BI. Essas plataformas são: Amazon Web Services (AWS), Microsoft Azure e Google Cloud Platform (GCP). As três plataformas compreendem modelos de Machine Learning e serviços que executam os processos necessários para a entrega de dados de um sistema de Business Intelligence.

A plataforma da Amazon [9] possui vários serviços que servem para a criação de pipelines[5] e entregam os dados preparados para a criação de modelos de machine learning (Figura 9). A partir dos serviços dela é possível estabelecer conexão com as fontes de dados, catalogar as informações que serão utilizadas, desenvolver o ETL e gerar modelos.

A Figura 8.1 mostra serviços da AWS que são capazes de armazenar os dados das origens (Amazon Redshift, Amazon S3, Amazon RDS, Amazon EC2, Amazon Kinesis, Amazon MSK), catalogar os dados (AWS Glue Data Catalog), desenvolver ETL (AWS Glue Studio), automatizar o processo de carga e transformação dos dados (AWS Glue

[5] Conjunto de processos capazes de entregar os dados prontos.

ETL), gerar insights a partir dos dados e criar modelos de machine learning (Amazon Athena, Amazon Redshift, Amazon EMR, Amazon Sagemaker, Amazon QuickSight). Todos os serviços mencionados podem ser utilizados para o desenvolvimento de uma solução de BI.

Figura 8.1: AWS. AWS Glue.

Disponível em <https://aws.amazon.com/pt/glue/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc>. Acesso em Abril de 2022.

A arquitetura da Microsoft Azure [10] possui a mesma estrutura que a arquitetura da AWS (Figura 8.2). Ela possui serviços para armazenamento de dados da fonte (Azure Data Lake Storage), para criação de pipelines (Azure Synapse Analytics) e para criação, preparação e treino de modelos de machine learning (Azure Synapse Analytics).

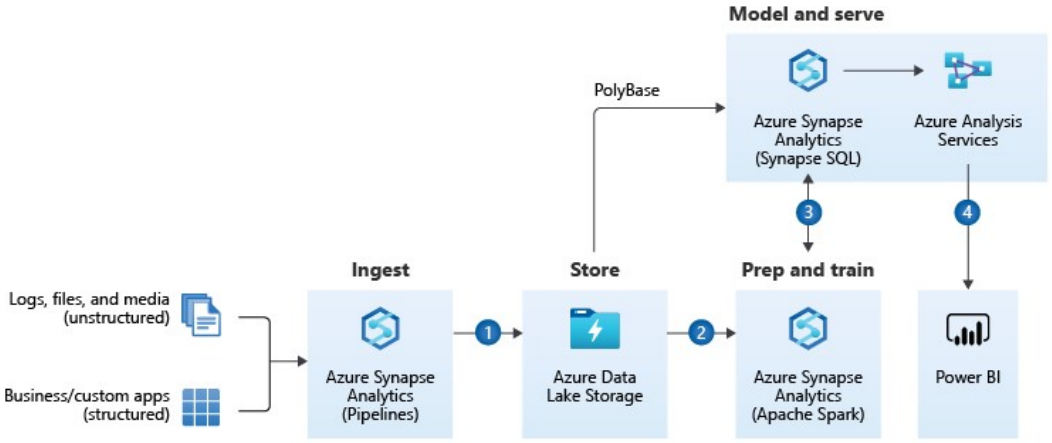


Figura 8.2: Microsoft. Enterprise Data Warehouse.

Disponível em <https://docs.microsoft.com/pt-br/azure/architecture/solution-ideas/articles/enterprise-data-warehouse>. Acesso em Abril de 2022.

A arquitetura da Google [11] segue o mesmo padrão das arquiteturas mostradas anteriormente (Figura 8.3). O serviço de armazenamento dos dados que foram extraídos da fonte de dados é o Cloud Storage. O BigQuery ou o Cloud SQL são os serviços para o desenvolvimento do ETL. O Cloud Dataflow e Cloud Pub/Sub servem para automatizar o processo de ETL. No Google Cloud Platform também é possível desenvolver os modelos de machine learning dentro do BigQuery.

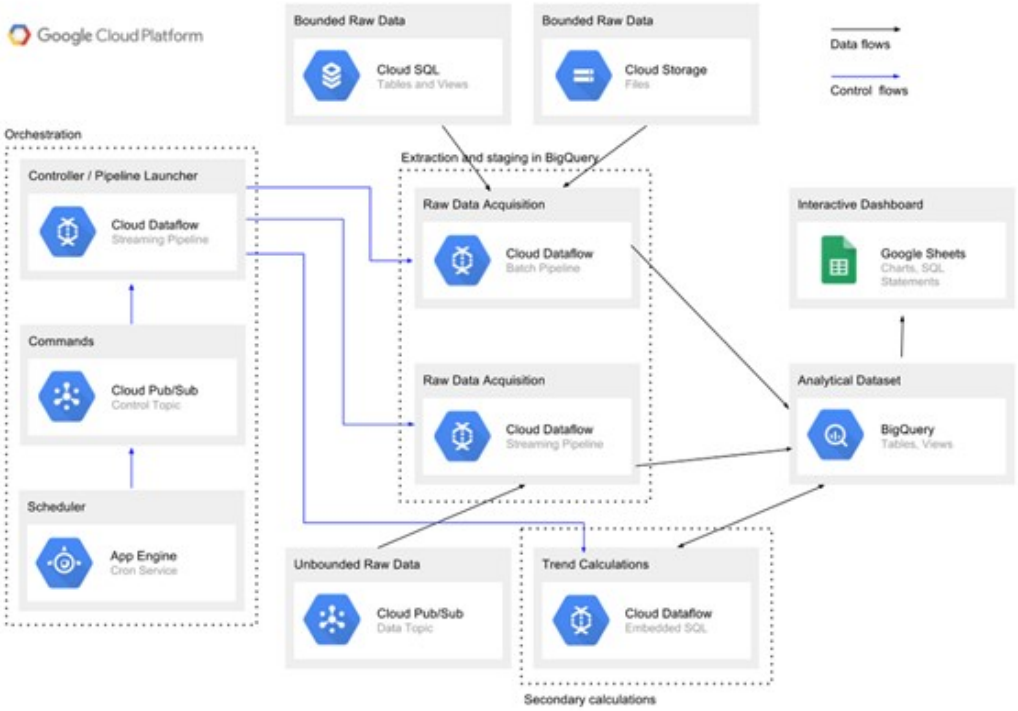


Figura 8.3: Google Cloud. Designing ETL architecture for a cloud-native data warehouse on Google Cloud Platform.

Disponível em <https://cloud.google.com/blog/products/bigquery/designing-etl-architecture-for-a-cloud-native-data-warehouse-on-google-cloud-platform>. Acesso em Abril de 2022.

Logo após avaliar as plataformas da Amazon, Microsoft e Google, procurei o relatório mais recente do quadrante mágico Gartner [2], que consiste numa representação gráfica de como estão performando as ferramentas de Business Intelligence no mercado. Avaliei o quadrante mágico para complementar os estudos das plataformas citadas acima e a partir disso, encontrei o Quadrante Mágico da Gartner de 2022 para plataformas de análise dos dados e Business Intelligence (Figura 8.4) e nele consta o Google classificado com uma ferramenta desafiadora.

Figura 8.4: Qlik. Quadrante Mágico™ do Gartner® 2022 para Plataformas de Analytics e Business Intelligence.

Disponível em<https://www.qlik.com/pt-br/gartner-magic-quadrant-business-intelligence>. Acesso em Abril de 2022.

Além da classificação do Quadrante Mágico ter ajudado na escolha da plataforma a ser utilizada para o tratamento dos dados, tenho maior familiaridade com o Google Cloud Platform, portanto essa foi a ferramenta escolhida para o desenvolvimento.

Os serviços desta plataforma que foram utilizados para o tratamento dos dados foram o Google Cloud Storage [4] e o Google Cloud Bigquery [5].

De acordo com a documentação do Google, o Storage, também chamado de GCS, consiste em um serviço de file system para o armazenamento de objetos de todos os tamanhos e de vários tipos. O Bigquery é um serviço que serve para gerenciar e analisar dados com diversos recursos integrados, como o aprendizado de máquina.

A partir da escolha desses serviços, foi feito o tratamento dos dados existentes nos relatórios indicados na seção 5 - Levantamento e entendimento dos dados do projeto.

Inicialmente, todos os relatórios, tanto os relatórios simulados, quanto os relatórios fornecidos pela coordenação do DI, foram carregados no Storage. Como esse serviço que armazena arquivos possui integração com o Bigquery, foram criadas tabelas externas no Google Cloud Bigquery, ou seja, tabelas que são criadas a partir dos dados dos arquivos armazenados no GCS.

Após levar os dados das planilhas para o serviço de tratamento e análise de dados (Bigquery), foi possível realizar as transformações necessárias nos dados para o desenvolvimento dos modelos.

O relatório Vagas_Disciplinas foi estruturado com dados de cada período em abas diferentes no excel. No tratamento desses dados foi necessário unificar os dados de todas essas abas e criar uma coluna indicando o período referente àquele dado. Além disso, algumas colunas desse relatório foram desconsideradas nesse processo. As colunas que foram mantidas e utilizadas para análises do modelo foram período, disciplina, código da turma, vagas oferecidas e vagas ocupadas.

A coluna que foi criada para indicar o período é uma coluna de data. As datas referentes ao mês de Janeiro indicam os dados do primeiro período do ano. Já as datas referentes ao mês de Agosto indicam os dados do segundo período.

Além disso, foi criada a coluna t_periodo, como mostra a tabela 8.1, que enumera todas as linhas utilizadas do relatório. Essa coluna foi criada especificamente para o modelo de previsão de quantidade de vagas, que será explicado na seção 8 - Modelos de Machine Learning.

Campo	Definição
Período	Indica o código do período.
Disciplina	Indica o código da disciplina.
Código da Turma	Indica o código da turma.
Vagas oferecidas	Indica a quantidade de vagas oferecidas.
Vagas ocupadas	Indica a quantidade de vagas ocupadas.
t_periodo	Enumera as linhas utilizadas do relatório para o desenvolvimento do modelo.

Tabela 8.1: Descrição de dados após os tratamentos feitos no Google Cloud Bigquery.

Para a realização desses tratamentos, foi necessário criar outra tabela no Bigquery. Os dados originais dos relatórios são carregados em uma tabela externa e para o ajuste desses dados, é criada uma tabela com as regras de mudança dos dados aplicadas.

O único relatório que sofreu tratamento foi o de Vagas_Disciplinas. Os demais relatórios carregados na plataforma foram relatórios simulados por mim e, por isso, não houve a necessidade de alteração.

A partir desses tratamentos realizados no Bigquery via SQL, foi possível preparar o dado da melhor forma para a criação dos modelos.

9. Modelos de Machine Learning

O projeto de Sistema de Tomada de Decisão na Fase de Pré-matrícula em uma Universidade Particular tem como objetivo auxiliar a coordenação do Departamento de Informática da PUC-Rio na previsão da quantidade de vagas que será disponibilizada para os alunos em cada período e no remanejamento dos professores em cada turma.

A partir destes objetivos, foram realizados estudos para estabelecer os modelos de Machine Learning que serão desenvolvidos para apoiar a decisão do coordenador do departamento.

O primeiro modelo escolhido foi o ARIMA (Autoregressive Integrated Moving Average), que consiste em analisar dados que foram gerados no passado e avaliar padrões para prever o futuro.

O modelo ARIMA [6] foi inicialmente escolhido para solucionar o problema da previsão de quantidade de vagas, pois é um modelo de fácil implementação, que garante previsões probabilísticas e apresenta informações sobre as mudanças ao longo do tempo. Essas propriedades do modelo facilitariam o desenvolvimento da solução proposta para a questão da previsibilidade da quantidade de vagas de uma disciplina.

Após iniciar o desenvolvimento do modelo, constatei que os dados de quantidade de vagas são apresentados duas vezes ao ano, ou seja, no primeiro e no segundo período do ano.

Pelo fato de o ARIMA ser um modelo de série temporal que analisa dados diários, a performance do modelo a partir dos dados periódicos de quantidade de vagas não foi muito satisfatória. Analisando os dados de teste utilizados para o desenvolvimento da previsão e comparando-os com os dados gerados pelo modelo, a diferença entre as quantidades de vagas reais e previstas foi grande.

Mesmo com esse modelo tendo uma acurácia[6] média, ou seja, variando de 30 a 70%, foi decidido manter esses resultados e analisar outros modelos que poderiam ter uma melhor performance.

Após alguns estudos, foi considerado o desenvolvimento do modelo de Regressão Linear [7], que compreende a análise de métodos estatísticos para a previsão de valores das variáveis que compõem a análise.

Para o desenvolvimento desse modelo, eu carreguei as tabelas que foram tratadas no serviço Bigquery do Google Cloud, especificado na seção 7 - Ferramentas de Business Intelligence, em um notebook python (Figura 9.1) para a manipulação dos dados através de dataframes[7].

---

[6] É calculada através do erro entre o valor previsto e o valor real.
[7] Tabelas manipuláveis que possuem linhas(registros) e colunas(campos).

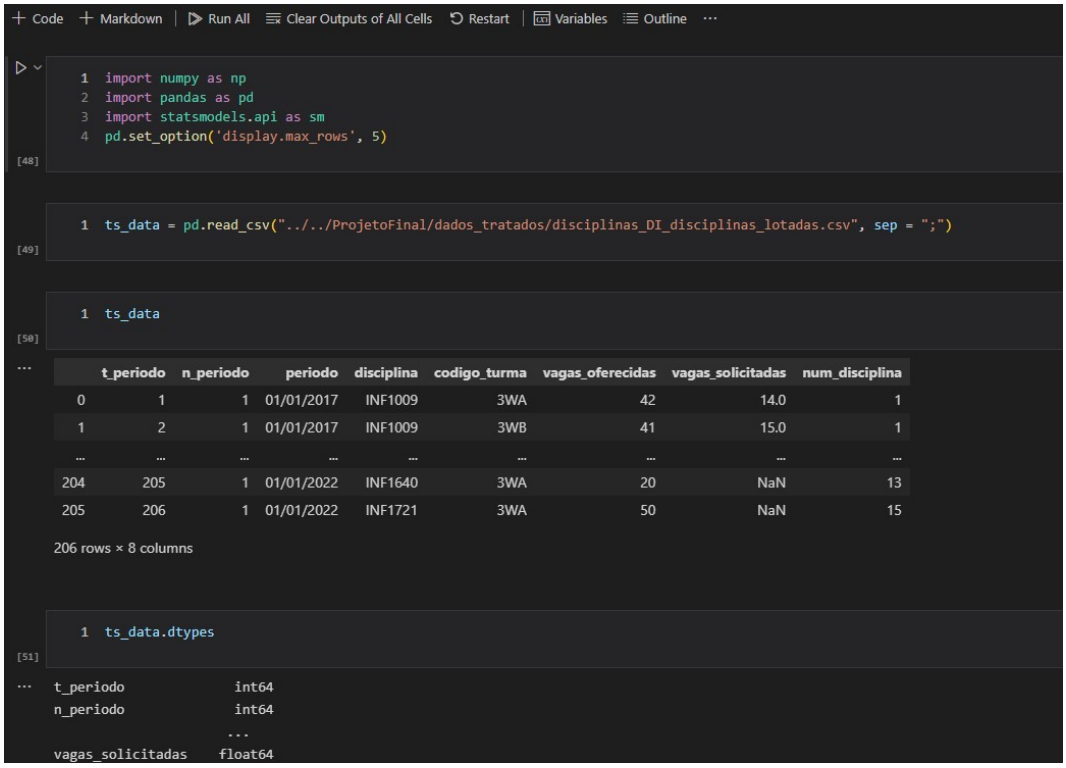


Figura 9.1: Notebook python utilizado por mim para o desenvolvimento dos modelos preditivos.

No notebook eu li o arquivo relativo às vagas das disciplinas como um dataframe e, com isso, eu iniciei o desenvolvimento do modelo.

A partir da coluna t_periodo que eu criei no tratamento dos dados, eu gerei duas novas colunas no dataframe chamadas t_periodo2 e t_periodo3 (Tabela 9.1) para criar um tempo linear elevando t_periodo2 ao quadrado e elevando t_periodo3 ao cubo. Essas colunas foram criadas com o intuito de ajudar a estimar os efeitos causados pelas mudanças periódicas do dado.

Campo	Definição
Período	Indica o código do período.
Disciplina	Indica o código da disciplina.
Código da Turma	Indica o código da turma.
Vagas oferecidas	Indica a quantidade de vagas oferecidas.
Vagas ocupadas	Indica a quantidade de vagas ocupadas.
t_periodo	Enumera as linhas utilizadas do relatório para o desenvolvimento do modelo.
t_periodo2	Indica o tempo linear de t_periodo2 ao quadrado.
t_periodo3	Indica o tempo linear de t_periodo3 ao cubo.

Tabela 9.1: Descrição de dados após manipulação do dataframe.

Em seguida, eu separei o dataframe em quatro partes. A primeira parte é o trecho do dataframe que foi utilizado para o treinamento, a segunda parte é o trecho de validação, a terceira é o de treinamento e validação e a quarta é o trecho de previsão. A partir disso, separei as variáveis x das variáveis y, removi a variável y do dataframe que possui a variável de entrada e apliquei isso para cada parte do dataframe.

Após essa preparação do conjunto de dados eu iniciei o modelo de Regressão Linear treinando-o com o trecho de treinamento e adicionando uma coluna de dados constantes de treino para o modelo produzir uma estimativa do coeficiente de regressão.

Com a aplicação do modelo no trecho de treinamento, foi possível avaliar os coeficientes dos dados de treino, mas para avaliar a qualidade de predição do modelo é necessário saber o y de vagas ocupadas e o y previsto para poder comparar os valores. Dessa forma, criei um dataframe explicitando esses dois valores e para avaliar a regressão eu usei as métricas R2, MAE e RMSE. Essas métricas são calculadas comparando o y de vagas ocupadas com o y previsto e indicam a qualidade estatística do dado.

Para poder chegar no melhor resultado do modelo, foi essencial avaliar o resultado dessas métricas para todos os tempos lineares criados. Logo, repeti o mesmo processo para t_periodo2 e t_periodo3 com cada trecho do dataframe e avaliei qual possuía o melhor resultado.

Depois de analisar as métricas para cada caso testado, pude concluir que o modelo performa melhor com t_periodo3, pois os valores são menores do que com t_periodo2. Desse modo, foi gerado o dataframe final com a quantidade de vagas previstas.

Para o desenvolvimento do modelo ARIMA foi necessário realizar os mesmos preparos dos dados do modelo de Regressão Linear. Após esses preparos, eu utilizei o auto_arima, que é uma função da biblioteca pmdarima que gera os valores para a criação do melhor modelo.

A partir dos valores resultantes do auto_arima, executei a função SARIMAX para gerar a previsão e, com isso, obtive o resultado da quantidade de vagas previstas.

Tanto o modelo ARIMA quanto o modelo de Regressão Linear foram executados duas vezes. A primeira vez foi para os dados do relatório Vagas_Disciplina para prever a quantidade de vagas para as disciplinas que não estava totalmente lotadas e a segunda vez foi para o relatório que eu simulei com números aleatórios de 1 a 50 para prever a quantidade de vagas das turmas que tiveram mais vagas solicitadas do que vagas oferecidas.

No relatório que eu simulei os dados, eu inseri a quantidade de vagas ocupadas com valores aleatórios de 1 a 50, pois esses eram aproximadamente os valores mínimos e máximos respectivamente da quantidade de vagas ocupadas real.

10. Alocação de recursos

Para o desenvolvimento da alocação de professores, foi estudado o modelo KNN (K Nearest Neighbor). Esse modelo havia sido escolhido inicialmente para este desenvolvimento, pois ele utiliza o cálculo da distância das classes para entender a conformidade dos registros.

O KNN [12] é um dos principais algoritmos para definir o rótulo dos dados. Ele utiliza o cálculo de distância das amostras para estabelecer o rótulo de um novo dado verificando a proximidade dele com as outras informações existentes. A partir das informações de alocação dos professores nos períodos passados, com o uso do modelo KNN seria possível prever o comportamento das alocações futuras e, com isso, sanar o problema do remanejamento dos professores em cada turma.

Após inúmeras análises dos dados que foram disponibilizados e simulados para a criação do segundo modelo, foi constatado que o algoritmo KNN não solucionaria o problema da alocação dos professores nas turmas. Com isso, foram estudados algoritmos que teriam características específicas para resolver a questão do modelo.

Um dos algoritmos estudados foi o Gale-Shapley [13], que consiste em um algoritmo que casa os pares de elementos existentes sem deixar nenhum elemento de fora.

O algoritmo Gale-Shapley foi estudado, pois encontrei o artigo Student Course Allocation with Constraints [14], que mostrava um cenário parecido com o cenário da alocação de professores que precisava ser solucionado.

Outro algoritmo que também foi analisado para esse desenvolvimento foi o algoritmo Húngaro [15], que na verdade consiste em um método de otimização que utiliza a manipulação de matrizes.

O estudo desses dois algoritmos foi essencial para o entendimento de como desenvolver a solução do remanejamento dos professores. A partir disso, utilizando o notebook python gerei o código para resolver a questão da alocação e é possível entendê-lo com base no pseudocódigo (Figura 10.1) apresentado a seguir.

```
transforma professor_horarios em dataframe
transforma disciplina_horario em dataframe
transforma professor_disciplina em dataframe
transforma qtd_disciplinas_previsao em dataframe

Aloca_Professores()
    transforma professor_horarios em lista
    transforma disciplina_horario em lista
    transforma professor_disciplina em lista
    transforma qtd_disciplinas_previsao em lista

    loop no tamanho da lista professor_disciplina
        loop no tamanho da lista qtd_disciplinas_previsao
            se o codigo da disciplina de professor_disciplina for igual ao codigo de qtd_disciplinas_previsao, entao
                quantidade_turmas vai ser igual a quantidade de turmas de qtd_disciplinas_previsao
        enquanto quantidade_turmas for maior que zero, faça
            loop no tamanho da lista professor_horarios
                se o nome do professor de professor_disciplina for igual a professor_horarios, entao
                    se essa disciplina nao foi alocada, entao
                        indica que que essa disciplina foi alocada
                        loop no tamanho da lista disciplina_horario
                            se o horario e o dia da semana de professor_horarios e disciplina_horario for igual, entao
                                cria uma lista vazia lista_final
                                insere em lista_final o codigo, nome da disciplina, horario, dia da semana e nome do professor
    cria uma lista vazia disciplinas_resolvidas
    loop no tamanho da lista disciplina_horario
        se a disciplina de professor_disciplina existe em disciplinas_resolvidas, entao
        insere em disciplinas_resolvidas o dado da disciplina
        loop no tamanho da lista lista_final
            se a disciplina de lista_final for igual a disciplina de disciplina_horario, entao
                soma 1
        loop no tamanho de qtd_disciplinas_previsao
            se a disciplina de disciplina_horario for igual a qtd_disciplinas_previsao, entao
                qtd_turmas_verifica é igual a quantidade de turmas de qtd_disciplinas_previsao
        se soma for menor que qtd_turmas_verifica, entao
            qtd_diferenca vai ser igual a qtd_turmas_verifica menos soma
            loop no tamanho de qtd_diferenca
                loop no tamanho da lista professor_disciplina
                    se a disciplina de professor_disciplina for igual a disciplina_horario, entao
                        insere em lista_final o codigo, nome da disciplina, "?", "?", "?"
    retorna lista_final
```

Figura 10.1: Pseudocódigo da função Aloca_Professores

Inicialmente, os dados dos horários dos professores, dos horários das disciplinas, das disciplinas que os professores lecionam e da quantidade de turmas previstas no modelo de regressão linear são transformados em dataframes para poderem ser manipulados. Esses dataframes são chamados de professor_horarios, disciplina_horario, professor_disciplina e qtd_disciplinas_previsao.

Na função Aloca_Professores, os dataframes com os dados necessários para a alocação são transformados em listas com a função tolist().

Em seguida as listas sao percorridas para verificar se o nome do professor da lista de professor_disciplina existe na lista professor_horario e se a disciplina não foi alocada ainda.

Além disso, o horário e o dia da semana da lista professor_horarios é comparado com disciplina_horario para ver se é compatível.

Se essas informações forem correspondentes, então os dados de código da disciplina, nome da disciplina, horario, dia da semana e nome do professor são inseridos na lista final.

Após todos os loops terminarem, existe uma verificação das disciplinas que foram resolvidas. Se por acaso alguma disciplina não tiver professor alocado na lista final, então o horario, dia da semana e nome do professor são preenchidos com ponto de interrogação.

Depois que a função Aloca_Professores termina de ser executada, a função Calcula_Custo(Figura 10.2) é chamada.

```
Calcula_Custo()
    loop da quantidade de cenarios que serao criados
        gera cenarios com ordem aleatória do dataframe professor_disciplina
    chama a funcao Aloca_Professores para gerar a alocacao com esses cenarios
    loop na lista final gerada pela funcao Aloca_Professores
        insere em custo a quantidade de professores que nao foram alocados
        verifica os dois cenarios com os menores custos
            insere em lista_custo os dados dos cenarios com menores custos
    retorna lista_custo
```

Figura 10.2: Pseudocódigo da função Calcula_Custo.

Essa função gera vários cenários alterando a ordem dos dados do dataframe professor_disciplina. Com isso, ela chama a função Aloca_Professores para alocar os cenários gerados.

Ela verifica cada lista final gerada pela função de alocação e insere na variável custo a quantidade de professores que não foram alocados.

As duas listas que tiverem o menor custo são retornadas pela função que calcula o custo.

11. Visualização

Além de ser necessário avaliar ferramentas para o processo de extração, transformação e carga dos dados em um sistema de BI, também é indispensável analisar softwares de visualização de dados para a entrega final das informações que foram geradas para os usuários. Alguns dos softwares que são capazes de gerar as análises para a tomada de decisão são Microsoft Power BI e Tableau. Eles são capazes de acessar arquivos de diferentes formatos ou gerar conexões por meio de conectores com as ferramentas que preparam os dados para serem consumidos em um sistema de Business Intelligence.

A partir da minha maior familiaridade com a ferramenta Power BI e a boa conectividade dela com o serviço Google Cloud Bigquery, serviço escolhido para a realização do ETL do projeto, ela foi a ferramenta escolhida para o desenvolvimento do dashboard para os coordenadores dos cursos do Departamento de Informática.

O arquivo Power BI gerado possui três abas. A primeira aba possui o nome “Modelo Preditivo Regressão”, a segunda possui o nome “Modelo Preditivo Arima” e a terceira aba “Alocação de Professores”, como mostra a Figura 11.1.

Figura 11.1: Abas do dashboard.

A primeira aba do dashboard possui os resultados do modelo de regressão linear. Nessa visualização (Figura 11.2 e 11.3) é possível analisar três indicadores: “Nível de acerto do modelo”, “Média do Valor Previsto” e “Média do Erro Absoluto”.

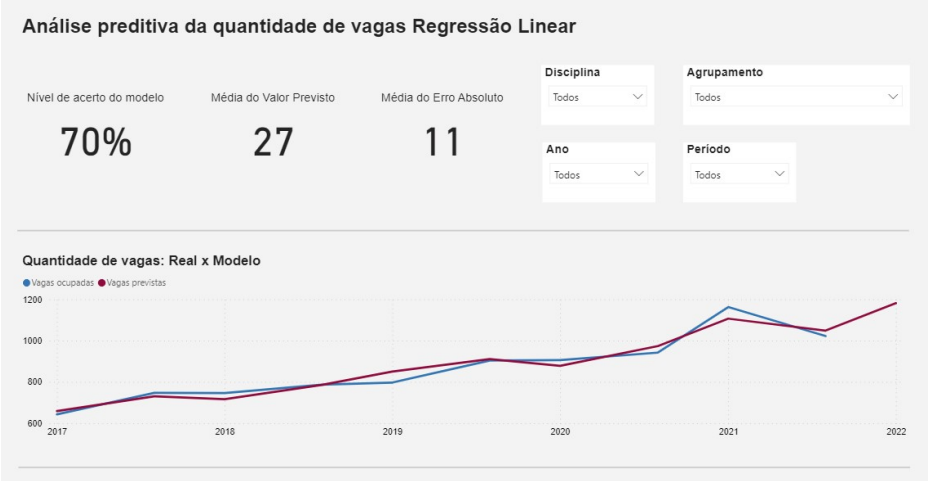
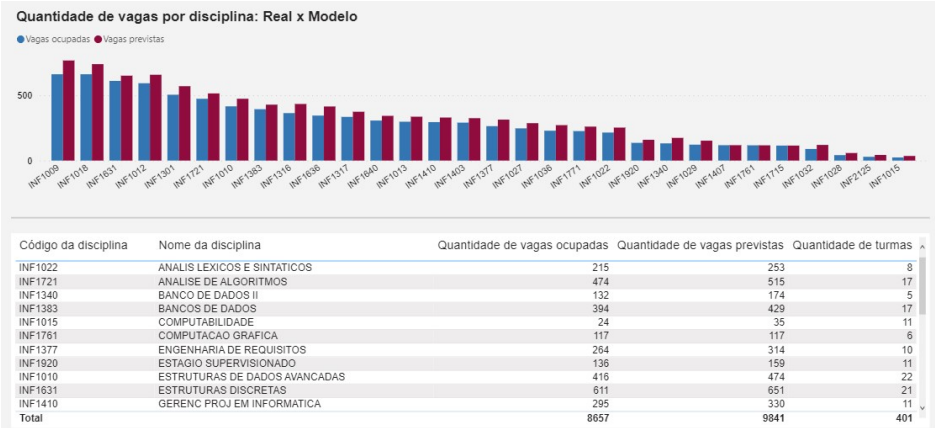


Figura 11.2: Primeira parte da visualização do modelo preditivo de regressão.



Código da disciplina	Nome da disciplina	Quantidade de vagas ocupadas	Quantidade de vagas previstas	Quantidade de turmas
INF1022	ANALIS LEXICOS E SINTATICOS	215	253	8
INF1721	ANALISE DE ALGORITMOS	474	515	17
INF1340	BANCO DE DADOS II	132	174	5
INF1383	BANCOS DE DADOS	394	429	17
INF1015	COMPUTABILIDADE	24	35	11
INF1761	COMPUTACAO GRAFICA	117	117	6
INF1377	ENGENHARIA DE REQUISITOS	264	314	10
INF1920	ESTAGIO SUPERVISIONADO	136	159	11
INF1010	ESTRUTURAS DE DADOS AVANCADAS	416	474	22
INF1631	ESTRUTURAS DISCRETAS	611	651	21
INF1410	GERENC PROJ EM INFORMATICA	295	330	11
Total		**8657**	**9841**	**401**

Figura 11.3: Segunda parte da visualização do modelo preditivo de regressão.

O primeiro indicador é a porcentagem do acerto do modelo calculado a partir da divisão do número de vagas ocupadas pelo número de vagas previstas.

O segundo indicador calcula a média do valor previsto gerado e o terceiro indicador é calculado através da diferença das vagas ocupadas pelas vagas previstas.

Além disso, essa visualização possui os filtros de disciplina, ano, período e agrupamento que podem ser selecionados pelo usuário e eles filtrarão toda a página.

O filtro de agrupamento foi gerado com o objetivo de agrupar os dados nos gráficos a partir das classificações. Essas classificações foram geradas por mim, mas podem ser alteradas pelos usuários clicando no campo indicado.

Abaixo dos indicadores e dos filtros, encontra-se o gráfico de linhas contendo a quantidade de vagas real e a quantidade de vagas do modelo no eixo y e o ano no eixo x. Nesse gráfico é possível analisar a performance desses dois valores ao longo do tempo.

Em seguida, tem o gráfico de barras da quantidade de vagas real e a quantidade de vagas do modelo no eixo y e as disciplinas no eixo x.

Nos dois gráficos é possível verificar a quantidade de turmas calculadas ao passar o mouse por cima deles.

Essa aba também contém uma tabela com as colunas código da disciplina, nome da disciplina, quantidade de vagas ocupadas, quantidade de vagas previstas e quantidade de turmas. Essa tabela possibilita verificar esses resultados juntos.

A aba do modelo preditivo arima (Figura 11.4 e Figura 11.5) possui os mesmos gráficos, filtros, indicadores e tabela que a aba do modelo preditivo regressão. A única diferença entre essas duas abas são as cores e os dados que foram utilizados. Em “Modelo Preditivo Arima” são usados os dados gerados pelo modelo ARIMA. Já em “Modelo Preditivo Regressão” são usados os dados do modelo de Regressão Linear.

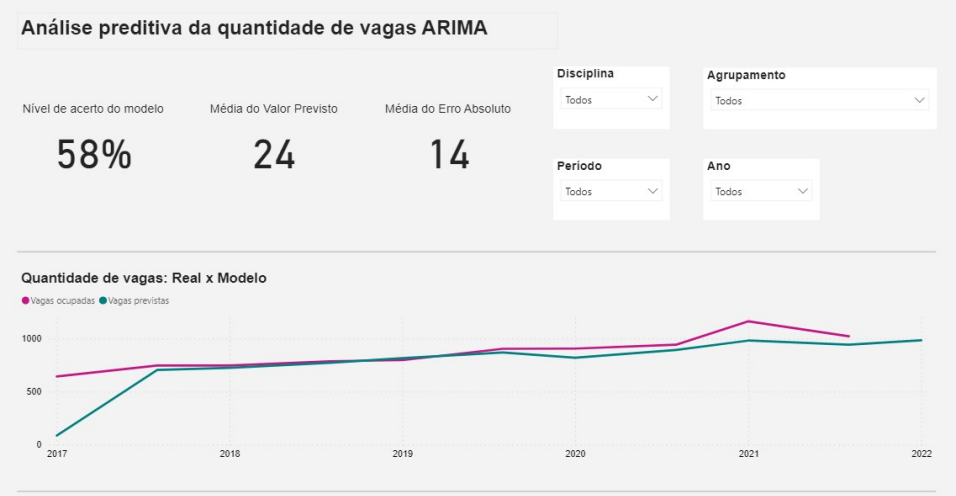
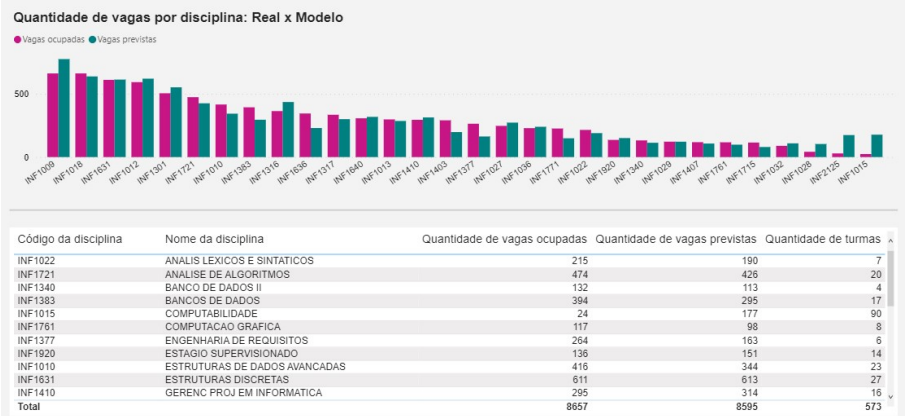


Figura 11.4: Primeira parte da visualização do modelo preditivo arima.



Código da disciplina	Nome da disciplina	Quantidade de vagas ocupadas	Quantidade de vagas previstas	Quantidade de turmas
INF1022	ANALIS LEXICOS E SINTATICOS	215	190	7
INF1721	ANALISE DE ALGORITMOS	474	426	20
INF1340	BANCO DE DADOS II	132	113	4
INF1383	BANCOS DE DADOS	394	295	17
INF1015	COMPUTABILIDADE	24	177	90
INF1761	COMPUTACAO GRAFICA	117	98	8
INF1377	ENGENHARIA DE REQUISITOS	264	163	6
INF1920	ESTAGIO SUPERVISIONADO	136	151	14
INF1010	ESTRUTURAS DE DADOS AVANCADAS	416	344	23
INF1631	ESTRUTURAS DISCRETAS	611	613	27
INF1410	GERENC PROJ EM INFORMATICA	295	314	16
Total		8657	8595	573

Figura 11.5: Segunda parte da visualização do modelo preditivo arima.

Além das abas de modelo, o dashboard possui a aba de alocação dos professores (Figura 11.6).

Código da disciplina	Nome da disciplina	Dia da semana	Horário	Nome do professor	Cenário
INF1009	LOGICA PARA COMPUTACAO	3a 5a	13 as 15	CECILIA REIS ENGLANDER LUSTOSA	1
INF1009	LOGICA PARA COMPUTACAO	3a 5a	15 as 17	CECILIA REIS ENGLANDER LUSTOSA	1
INF1009	LOGICA PARA COMPUTACAO	3a 5a	15 as 17	EDWARD HERMANN HAEUSLER	1
INF1010	ESTRUTURAS DE DADOS AVANCADAS	2a 4a	13 as 15	AUGUSTO CESAR ESPINDOLA BAFFA	1
INF1010	ESTRUTURAS DE DADOS AVANCADAS	2a 4a	15 as 17	LUIZ FERNANDO BESSA SEIBEL	1
INF1012	MODELAGEM DE DADOS	6a	13 as 15	IVAN MATHIAS FILHO	1
INF1012	MODELAGEM DE DADOS	6a	7 as 9	IVAN MATHIAS FILHO	1
INF1015	COMPUTABILIDADE	3a 5a	13 as 15	EDWARD HERMANN HAEUSLER	1
INF1022	ANALIS LEXICOS E SINTATICOS	2a 4a	11 as 13	EDWARD HERMANN HAEUSLER	1
INF1027	TESTE E MEDICAO DE SOFTWARE	2a 4a	11 as 13	MARCOS KALINOWSKI	1
INF1028	PROJETO E CONSTRUCAO SISTEMAS	3a 5a	15 as 17	ALESSANDRO FABRICIO GARCIA	1
INF1029	INT ARQUITETURA COMPUTADORES	3a 5a	9 as 11	ANDERSON OLIVEIRA DA SILVA	1
INF1032	INTRODUCAO A CIENCIA DOS DADOS	5a	15 as 18	HELIO CORTES VIEIRA LOPES	1
INF1036	PROBABILIDADE COMPUTACIONAL	3a 5a	7 as 9	ANA CAROLINA LETICHEVSKY	1
INF1301	PROGRAMACAO MODULAR	2a 4a	17 as 19	FLAVIO HELENO BEVILACQUA E SILVA	1
INF1301	PROGRAMACAO MODULAR	3a 5a	7 as 9	FLAVIO HELENO BEVILACQUA E SILVA	1
INF1316	SISTEMAS OPERACIONAIS	3a 5a	13 as 15	LUIZ FERNANDO BESSA SEIBEL	1
INF1316	SISTEMAS OPERACIONAIS	3a 5a	15 as 17	LUIZ FERNANDO BESSA SEIBEL	1
INF1316	SISTEMAS OPERACIONAIS	3a 5a	9 as 11	LUIZ FERNANDO BESSA SEIBEL	1
INF1317	REDES DE COMPUTADORES	3a 5a	9 as 11	SERGIO COLCHER	1
INF1340	BANCO DE DADOS II	3a 5a	7 as 9	MARCOS VIANNA VILLAS	1
INF1377	ENGENHARIA DE REQUISITOS	3a 5a	15 as 17	JULIO CESAR SAMPAIO DO PRADO LEITE	1
INF1377	ENGENHARIA DE REQUISITOS	3a 5a	17 as 19	EDMUNDO BASTOS TORREAO	1
INF1383	BANCOS DE DADOS	3a 5a	15 as 17	SERGIO LIFSCHITZ	1
INF1403	INTR INT HUMANO-COMPUTADOR	3a 5a	11 as 13	SIMONE DINIZ JUNQUEIRA BARBOSA	1
INF1410	GERENC PROJ EM INFORMATICA	2a 4a	15 as 17	MARCANTONIO GIUSEPPE MARIA C FABRA	1
INF1631	ESTRUTURAS DISCRETAS	3a 5a	13 as 15	EDUARDO SANY LABER	1
INF1636	PROGRAMACAO ORIENTADA OBJETOS	2a 4a	7 as 9	IVAN MATHIAS FILHO	1
INF1636	PROGRAMACAO ORIENTADA OBJETOS	3a 5a	13 as 15	IVAN MATHIAS FILHO	1
INF1636	PROGRAMACAO ORIENTADA OBJETOS	3a 5a	7 as 9	IVAN MATHIAS FILHO	1
INF1636	PROGRAMACAO ORIENTADA OBJETOS	6a	15 as 17	IVAN MATHIAS FILHO	1
INF1640	REDES DE COMUNICACAO DE DADOS	?	?	?	1

Figura 11.6: Visualização da alocação dos professores.

Essa aba possui os filtros de cenário, disciplina, nome do professor e agrupamento.

O usuário pode filtrar o dado que desejar e isso irá alterar a visualização da tabela que se encontra ao lado dos filtros.

A partir disso, os coordenadores do Departamento de Informática são capazes de analisar e tomar decisão utilizando esses resultados apresentados nas visualizações.

12. Apresentação dos resultados para os coordenadores

Na definição da metodologia deste projeto, a etapa de avaliação dos coordenadores do Departamento de Informática, Ivan Mathias Filho e Augusto Baffa, foi considerada importante, pois no desenvolvimento de um projeto é importante ter a opinião do usuário.

Com isso, foi agendada uma reunião via Zoom[8] com cada um dos coordenadores para explicar o conceito deste projeto e apresentar o dashboard.

O professor Ivan explicou a importância de ter os dados de quantos alunos estão aptos a se inscrever nas disciplinas e quantos alunos foram reprovados ou trancaram a matéria, pois isso mostra o fluxo de alunos para o período seguinte. Esses dados não existem na PUC-Rio, mas seriam interessantes para o projeto.

Ele comentou sobre a complexidade da alocação das salas para cada turma e de como é feita hoje a previsão da quantidade de vagas por disciplina e a alocação dos professores. Por conta disso, ele elogiou a visualização dos dados, pois ela facilitaria esses dois processos.

Além disso, ele comentou sobre a importância de separar os períodos de cada ano e sugeriu um “fluxo” para trabalhos futuros que consiste em avaliar aspecto dinâmico, represamentos e os pré-requisitos das disciplinas.

O professor Baffa falou da importância da quantidade das turmas, porque o espaço é muito concorrido pelos alunos e, assim como o professor Ivan, mencionou a importância da alocação das salas.

Ele também explicou que o ARIMA, por ser um modelo de série temporal, a performance não foi tão boa quanto o modelo de regressão linear e que o modelo de regressão linear seria o melhor modelo para o desenvolvimento da previsão da quantidade de vagas por disciplina.

Baffa indicou que seria interessante apresentar a margem do canal de confiança (bollinger channel), que é um indicador que mostra uma linha média móvel central mais linhas de canal a uma distância acima e abaixo. Ele disse que seria bom ter isso para não ter só o valor fixo calculado, mas também uma margem dos valores.

Ele comentou da questão dinâmica das ondas de presença e ausência de professores e que é desejável a integração dos dados entre os sistemas da PUC.

[8] Aplicativo de reuniões virtuais.

13. Trabalhos futuros

Para os trabalhos futuros deste projeto, é interessante realizar a integração dos dados entre os sistemas da PUC-Rio, como disse o professor Augusto Baffa na reunião com os coordenadores.

Armazenar os dados da simulação de matrícula e ter os dados de quantos alunos estão aptos a se inscrever nas disciplinas e quantos alunos foram reprovados ou trancaram a matéria seria muito importante para gerar melhorias no projeto.

Automatizar os processos extração, transformação e carga dos dados é uma melhoria que facilitaria e otimizaria a integração entre as ferramentas utilizadas.

Referências Bibliográficas

[1] Oracle Brasil. O que é ETL? Disponível em <https://www.oracle.com/br/integration/what-is-etl/>. Acesso em Agosto de 2022.

[2] Data Skills. Business Intelligence Methodology. Disponível em <https://www.dataskills.ai/business-intelligence-methodology/#gref>. Acesso em Abril de 2022.

[3] Woebcken, Cayo. Quadrante mágico Gartner: o que é e qual a aplicabilidade? Disponível em <https://rockcontent.com/br/blog/quadrante-magico-gartner/>. Acesso em Agosto de 2022.

[4] Google Cloud. Cloud Storage. Disponível em < https://cloud.google.com/storage/?utm_source=google&utm_medium=cpc&utm_campaign=latam-BR-all-pt-dr-BKWS-all-all-trial-e-dr-1011454-LUAC0008670&utm_content=text-ad-none-any-DEV_c-CRE_429691579825-ADGP_Hybrid%20%7C%20BKWS%20-%20EXA%20%7C%20Txt%20~%20Storage_Cloud%20Storage-KWID_43700040369789728-kwd-308056723381&utm_term=KW_google%20cloud%20storage-ST_Google%20Cloud%20Storage&gclid=CjwKCAjw7p6aBhBiEiwA83fGuksjAtg-QoyALXAlk3qsSm8RuyLt4MLkRcL1pSAYTS9E7BTaMrPbvhoCVAYQAvD_BwE&gclsrc=aw.ds>. Acesso em Outubro de 2022.

[5] Google Cloud. Bigquery. Disponível em <https://cloud.google.com/bigquery/docs/introduction?hl=pt-br>. Acesso em Outubro de 2022.

[6] E. P. LIMA1 , S. A. POZZA2 , M. L. GIMENES3 , J. R. COURY2. USO DE MODELOS ARIMA SAZONAIS NO ESTUDO DA SÉRIE TEMPORAL DE MP10 DA CIDADE DE SÃO CARLOS. Disponível em <https://www.researchgate.net/profile/Simone-Pozza/publication/260517612_USO_DE_MODELOS_ARIMA_SAZONAIS_NO_ESTUDO_DA_SERIE_TEMPORAL_DE_MP_10_DA_CIDADE_DE_SAO_CARLOS/links/00b7d5317d17d4cbd2000000/USO-DE-MODELOS-ARIMA-SAZONAIS-NO-ESTUDO-DA-SERIE-TEMPORAL-DE-MP-10-DA-CIDADE-DE-SAO-CARLOS.pdf>. Acesso em Outubro de 2022.

[7] Antunes Rodrigues, Sandra Cristina. Modelo de Regressão Linear e suas Aplicações. Disponível em <https://www.proquest.com/openview/ddde3bb8c207d5a138545c2fcd34a6ce/1?pq-origsite=gscholar&cbl=2026366&diss=y>. Acesso em Outubro de 2022.

[8] Red Hat. O que é um provedor de serviços em nuvem? Disponível em <https://www.redhat.com/pt-br/topics/cloud-computing/what-are-cloud-providers>. Acesso em Junho de 2022.

[9] AWS. AWS Glue. Integração de dados simples, escalável e sem servidor. Disponível em <https://aws.amazon.com/pt/glue/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc>. Acesso em Junho de 2022.

[10] Microsoft. Enterprise Data Warehouse. Disponível em <https://learn.microsoft.com/pt-br/azure/architecture/solution-ideas/articles/enterprise-data-warehouse>. Acesso em Junho de 2022.

[11] Google Cloud. Designing ETL architecture for a cloud-native data warehouse on Google Cloud Platform. <Disponível em <https://cloud.google.com/blog/products/bigquery/designing-etl-architecture-for-a-cloud-native-data-warehouse-on-google-cloud-platform>. Acesso em Junho de 2022.

[12] Inferir. Algoritmo KNN para classificação. Disponível em <https://inferir.com.br/artigos/algoritimo-knn-para-classificacao/>. Acesso em Junho de 2022.

[13] Akinobu Eguchi1, Satoru Fujishige2, and Akihisa Tamura2. A Generalized Gale-Shapley Algorithm for a Discrete-Concave Stable-Marriage Model. Disponível em <https://link.springer.com/chapter/10.1007/978-3-540-24587-2_51>. Acesso em Outubro de 2022.

[14] Akshay Utture2, Vedant Somani1, Prem Krishnaa1, and Meghana Nasre1. Student Course Allocation with Constraints. Disponível em <http://web.cs.ucla.edu/~akshayutture/papers/sea19Preprint.pdf>. Acesso em Agosto de 2022.

[15] João Cesar Guirado. O MÉTODO HÚNGARO PARA RESOLUÇÃO DE PROBLEMAS DE OTIMIZAÇÃO. Disponível em <http://sbemparana.com.br/arquivos/anais/epremxii/ARQUIVOS/MINICURSOS/autores/MCA016.pdf>. Acesso em Agosto de 2022.