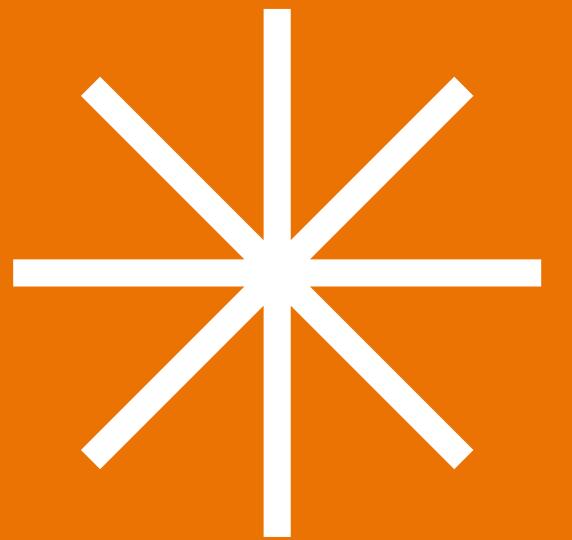




Case Analytics

Maycon Felipe Mota - Desenvolvedor (E)





Utilizando os dados enviados para o teste prático, realizei as análises utilizando o Python. Todos os códigos estão disponíveis no GitHub (<https://github.com/FelipeGaleao/case-gol-analytics>). Importante destacar que o link para extração dos microdados para voos se encontrava fora do ar, portanto, utilizei um alternativo que utilizava para fonte de estudos.

1ª QUESTÃO

Faça um ranking para o número total de PAX por dia da semana.

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

1ª Questão - Faça um ranking para o número total de PAX por dia da semana.

Código da resolução do problema



```
# Leitura do Dataset do Case 01
df_case01 = pd.read_excel('./input/case_analytics-ds (1).xlsx', sheet_name='Dados')

# Gerar coluna dia da semana
df_case01['Dia_Semana_Venda']= df_case01['Data Venda'].dt.day_name(locale='pt')

# Agregar pela soma, agrupando pelo dia da semana e criar um ranking do maior ao menor
df_ranking_pax = df_case01.groupby(by= df_case01['Dia_Semana_Venda']).sum()
df_ranking_pax['rank'] = df_ranking_pax['PAX'].rank(ascending=False)
df_ranking_pax = df_ranking_pax.sort_values(by= ['rank'])

# Exibir o DF com PAX, RPK e posição do ranking
df_ranking_pax[['PAX', 'RPK', 'rank']]
```

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

1ª Questão - Faça um ranking para o número total de PAX por dia da semana.

Quarta-Feira, Quinta-feira
e Sexta-feira.

são os dias da semana onde nós aproximamos mais pessoas.

	DIA	PAX	RPK
1	Quarta-feira	293025	26263705
2	Quinta-feira	258488	23501102
3	Sexta-feira	255625	23309936
4	Sábado	253467	23201577
5	Domingo	241620	22388695
6	Segunda-feira	236316	23275908
7	Terça-feira	219802	20366840

Figura 1. Resultado do dataframe após execução do código do slide anterior. (Fonte: Autor, 2021)

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

2ª Questão - Qual a correlação de sábado e domingo somados com o total de RPK?

Código da resolução do problema



```
# Para verificarmos a correlação de sábado e domingo somados com o valor total de RPK, irei criar um gráfico de barras.

# Transformar o dataset anterior e resetar o índice
df_correlacao = df_ranking_pax.reset_index()[['Dia_Semana_Venda', 'PAX', 'RPK']]

# Fazer uma linha somando sábado e domingo
df_correlacao = df_correlacao.append(df_correlacao.loc[(df_correlacao['Dia_Semana_Venda'] == 'Sábado') | (df_correlacao['Dia_Semana_Venda'] == 'Domingo')].sum(), ignore_index=True)

# Eliminar linha de sábado e domingo
df_correlacao.drop([0, 5])
```

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

2ª Questão - Qual a correlação de sábado e domingo somados com o total de RPK?

A questão não apresenta informações concretas para que seja possível criar uma correlação, uma vez que para que seja possível fazer correlações deve-se, ao mínimo, utilizar duas medidas sendo X e Y em uma escala contínua para que seja possível obter o coeficiente de correlação. Contudo, executei a análise fazendo a relação entre a soma do número de RPK de sábado e domingo com o total de RPKs, utilizando o DataFrame do Ranking Semanal.

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

2ª Questão - Qual a correlação de sábado e domingo somados com o total de RPK?

Código da resolução do problema



```
# Para verificarmos a correlação de sábado e domingo somados com o valor total de RPK, irei criar um gráfico de barras.  
# Transformar o dataset anterior e resetar o índice  
df_correlacao = df_ranking_pax.reset_index()[['Dia_Semana_Venda', 'PAX', 'RPK']]  
  
# Fazer uma linha somando sábado e domingo  
df_correlacao = df_correlacao.append(df_correlacao.loc[(df_correlacao['Dia_Semana_Venda'] == 'Sábado') | (df_correlacao['Dia_Semana_Venda'] == 'Domingo')].sum(), ignore_index=True)  
  
# Remover as linhas referentes ao Sábado e Domingo  
df_correlacao = df_correlacao.drop([3, 4])  
  
# Criar uma coluna de porcentagem em relação ao TOTAL RPK executado pelos dias das semanas  
df_correlacao['pct_total'] = (df_correlacao['RPK']/df_correlacao['RPK'].sum())*100
```

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

2ª Questão - Qual a correlação de sábado e domingo somados com o total de RPK?

Sábado e Domingo correspondem 28.08% da soma total do RPK.

	RPK	%
Sábado e Domingo	45590272	28.08
Quarta-feira	26263705	16.18
Quinta-feira	23501102	14.47
Sexta-feira	23309936	14.36
Segunda-feira	23275908	14.34
Terça-feira	20366840	12.54

Figura 2. Resultado do dataframe após execução do código do slide anterior. (Fonte: Autor, 2021)

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

3ª Questão - Qual a média de 'Monetário' por mês por Canal? E a mediana?

Código da resolução do problema



```
# Gerando uma coluna de Número de mês
df_case01['Número_Mes'] = df_case01['Data Venda'].dt.month

# Ordenando valores pela data venda
df_case01 = df_case01.sort_values(by= 'Data Venda', ascending=False)

# Agrupando os dados agregando média e mediana por canal de venda e mês
df_media = df_case01.groupby(by= ['Canal de Venda', 'Número_Mes'])['Monetário Vendido'].agg( [ 'mean' ,
'median'])

# renomeando o nome das colunas
df_media.columns = [ 'Média', 'Mediana']
df_media
```

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

3ª Questão - Qual a média de 'Monetário' por mês por Canal? E a mediana?

Canal de Venda	Mes_Ano	Média	Mediana
Porta a Porta	01/2017	31320.569487	28823.025305
	02/2017	31710.148886	29711.580146
	03/2017	31055.939762	30488.191027
	04/2017	33319.496263	32421.609418
	11/2016	34638.071796	30262.977968
	12/2016	28615.452500	25527.519295
TeleVenda	01/2017	32868.881429	19085.123238
	02/2017	32208.445818	18926.690232
	03/2017	32132.246712	18666.879969
	04/2017	35795.781915	28091.885619
	11/2016	30963.459349	18757.707785
	12/2016	28045.721461	17567.600938
Telégrafo	01/2017	4234.022567	3782.248868
	02/2017	4307.626273	4045.893131
	03/2017	4568.148531	4345.559552
	04/2017	3778.591835	3458.884666
	11/2016	4614.310076	4029.132131
	12/2016	5212.648055	4227.577948

A maior média de passageiros é através do canal de vendas Telefone, sendo em Abril a maior média, 35.795 mil passageiros.

Já o porta a porta atingiu a maior média de passageiros em Novembro, alcançando 34.638 passageiros.

O segmento mais crítico: Telégrafo, atingiu o seu pico mínimo de passageiros em Abril, em contra-mão ao TeleVenda.

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

3ª Questão - Qual a média de 'Monetário' por mês por Canal? E a mediana?

Canal de Venda	Mes_Ano	Média	Mediana
Porta a Porta	01/2017	31320.569487	28823.025305
	02/2017	31710.148886	29711.580146
	03/2017	31055.939762	30488.191027
	04/2017	33319.496263	32421.609418
	11/2016	34638.071796	30262.977968
	12/2016	28615.452500	25527.519295
	TeleVenda	32868.881429	19085.123238
		32208.445818	18926.690232
		32132.246712	18666.879969
		35795.781915	28091.885619
	11/2016	30963.459349	18757.707785
Telégrafo	12/2016	28045.721461	17567.600938
	01/2017	4234.022567	3782.248868
	02/2017	4307.626273	4045.893131
	03/2017	4568.148531	4345.559552
	04/2017	3778.591835	3458.884666
	11/2016	4614.310076	4029.132131
	12/2016	5212.648055	4227.577948

Figura 4. Execução do código anterior, mas ordenando pela Média. (Fonte: Autor, 2021)



Portanto, analisando o canal de venda mais crítico: Telégrafo, é possível notar uma queda de 27,52% entre o pico máximo (Dezembro/2016) e pico mínimo (Abril/2017). Em contra-partida, os canais Televenda e Porta-porta, registraram seu pico máximo em Abril de 2017. Portanto, levanta-se a hipótese de estar ocorrendo uma possível migração de clientes do meio de canal Telégrafo ao Porta a Porta e Televendas.

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

4^a Questão - Crie um forecast de PAX por 'Local de Venda' para os próximos 15 dias a contar da última data de venda. (Aqui a técnica é livre)

PROPHET

Para o desenvolvimento do forecast, utilizei o Prophet. Uma biblioteca que consegue analisar séries temporais onde há sazonalidade não-linear anualmente, semanalmente e até diariamente. Possui um bom comportamento quando envolve dados históricos ou eventos sazonais. Uma dos objetivos da biblioteca é automatizar o maior números de passos para criar um modelo, isso é, ele realiza o tratamento automático de outliers, dados faltantes e mudanças drásticas na série temporal.

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

4ª Questão - Crie um forecast de PAX por 'Local de Venda' para os próximos 15 dias a contar da última data de venda. (Aqui a técnica é livre)

Código da resolução do problema

```
● ● ●

# Para analisar essa série temporal, irei utilizar a biblioteca Prophet que foi desenvolvida pelo
Facebook e é utilizada em grande parte das aplicações.
### https://facebook.github.io/prophet/

from prophet import Prophet

# Transformação do dataset para o formato ds / y, conforme a documentação do Prophet indica

df_forecast = df_case01[['Data Venda', 'PAX']]
df_forecast.columns = ['ds', 'y']
df_forecast = df_forecast.groupby(by = 'ds').sum()
df_forecast = df_forecast.reset_index()
df_forecast

# Intanciação do modelo, aqui fiz a definição de alguns parametros como o modo da sazonalidade,
# sazonalidade anual, mensal ou diária. Em seguida, adicionei as sazonalidades mensais e trimestrais, por
# fim, passei a biblioteca os feriados do Brasil.

m = Prophet(seasonality_mode='additive', yearly_seasonality=False, weekly_seasonality=True,
            daily_seasonality=False, seasonality_prior_scale=10, changepoint_prior_scale=3,
            holidays_prior_scale=20)
m.add_seasonality(name='monthly', period=30.5, fourier_order=12)
m.add_seasonality(name='quartely', period=365.25/4, fourier_order=5, prior_scale=15)
m.add_country_holidays(country_name='BR')

# Ajuste do modelo para treinamento

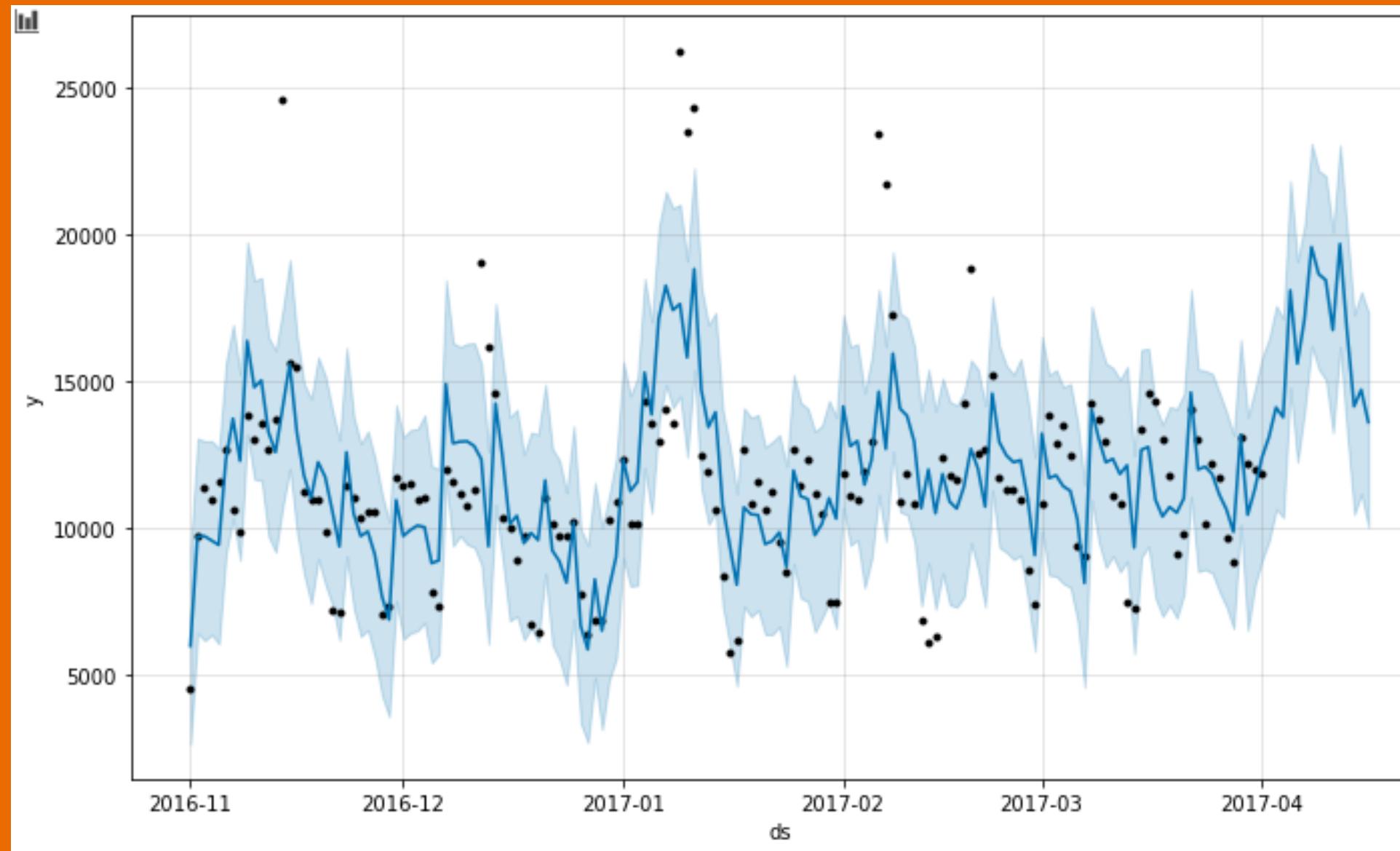
m.fit(df_forecast)
```

Inicialmente, gerei uma previsão para todos os canais de vendas para o próximo 15 dias para entender como o Prophet se comportaria.

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

4^a Questão - Crie um forecast de PAX por ‘Local de Venda’ para os próximos 15 dias a contar da última data de venda. (Aqui a técnica é livre)



Com base no dataset, o modelo sugeriu que nos próximos 15 dias, teremos um novo pico de vendas.

Atingindo seu ápice entre 10 a 14 de Abril.

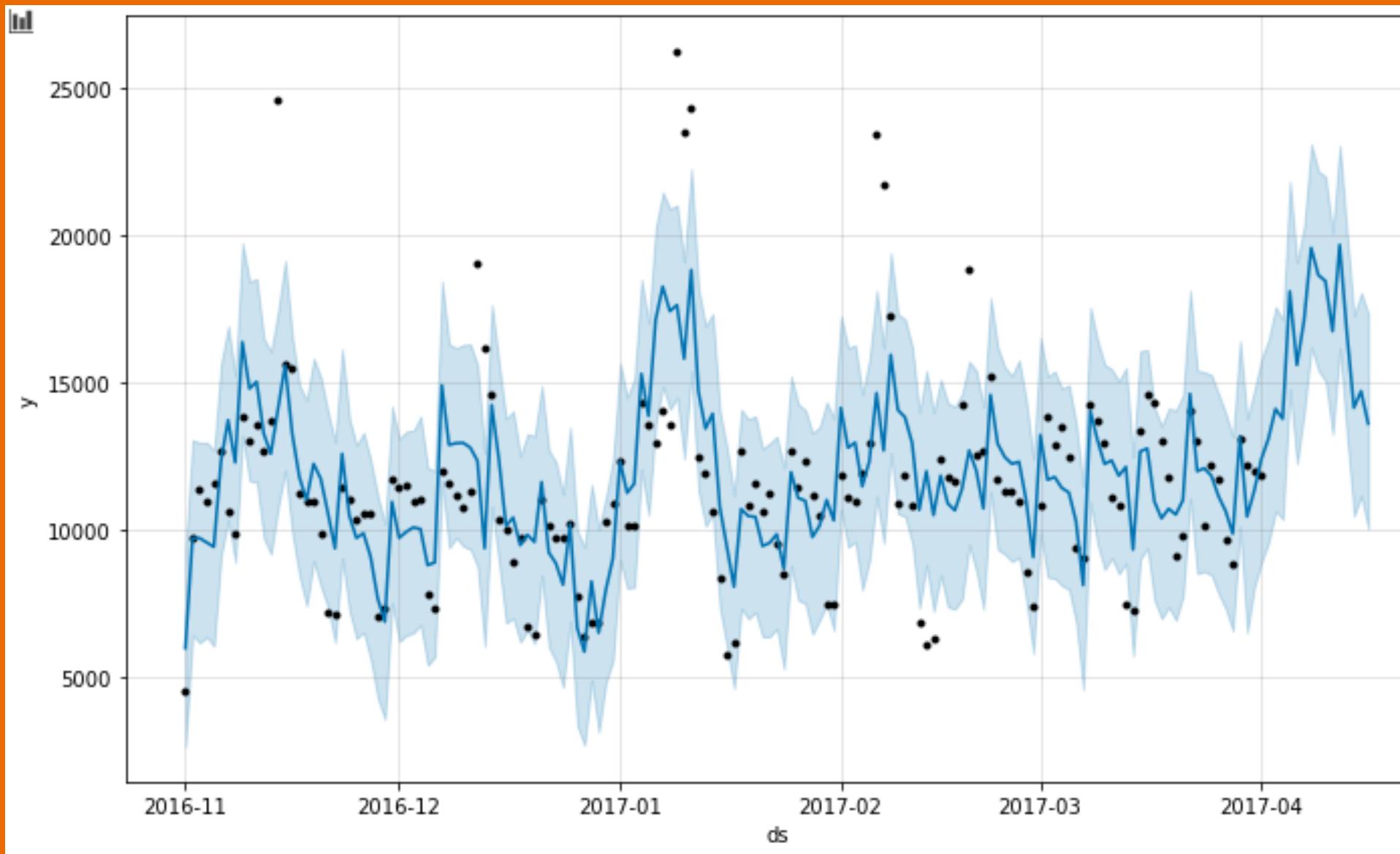
Os pontos pretos são os valores realizados e em Azul, o que o modelo conseguiu prever.

Figura 5. Modelo usado para realizar o Forecast, a linha em azul é o que o modelo prevê, e os pontos pretos são o realizado..
(Fonte: Autor, 2021)

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

4^a Questão - Crie um forecast de PAX por ‘Local de Venda’ para os próximos 15 dias a contar da última data de venda. (Aqui a técnica é livre)



Analisando cada um dos dados,
Os pontos pretos são os valores realizados e em Azul, o
que o modelo conseguiu prever.

Em testes, o presente modelo apresentou os seguintes
valores:

Figura 6. Modelo usado para realizar o Forecast, a linha em azul é o que o modelo prevê, e os pontos pretos são o realizado..
(Fonte: Autor, 2021)

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

5^a Questão - Supondo que você precisa gerar um estudo para a área responsável, com base em qualquer modelo ou premissa, qual ‘Local de Venda’ você considera mais crítico. Por quê?

Através do modelo gerado, foi gerado uma previsão aos próximos 15 dias. Após dia 02 de Abril, a companhia terá uma ascensão, onde entre dia 06 a 16 de Abril atingirá um pico de passageiros transportados. A empresa deve se preparar com maior número de recursos nos locais de vendas, sobretudo no Vast e Arena, onde aponta o maior pico de passageiros.



Figura 7. Forecast para os próximos 15 dias por local de venda. (Fonte: Autor, 2021)

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

6ª Questão - Criar modelo relacionando o comportamento de venda com variáveis não apresentada nos dados (Ex : PIB, Dolar, e etc)

Para o desenvolvimento do modelo de correlação, utilizei o conjunto de dados disponibilizados pela br.investing.com. Para realizar a correlação, utilizei uma janela de tempo de uma semana entre os dois conjuntos de dados.

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

6ª Questão - Criar modelo relacionando o comportamento de venda com variáveis não apresentada nos dados (Ex : PIB, Dolar, e etc)

Código da resolução do problema

```
## Leitura do Dataset de Dados Históricos sobre o dólar. Fonte: https://br.investing.com/cotacao\_dolar = pd.read\_csv\('./input/USD\_BRL Dados Históricos.csv'\)
## Transformação para parsear o campo de data
cotacao_dolar['Data'] = cotacao_dolar['Data'].str.replace('.', '/')
cotacao_dolar['Data'] = pd.to_datetime(cotacao_dolar['Data']) ## Parsear o campo data
cotacao_dolar['Último'] = pd.to_numeric(cotacao_dolar['Último'].str.replace(',', '.')) ## Parsear o campo Último e trocar , por .
cotacao_dolar['Máxima'] = pd.to_numeric(cotacao_dolar['Máxima'].str.replace(',', '.')) ## Parsear o campo Máxima e trocar , por .
cotacao_dolar['Var%'] = cotacao_dolar['Var%'].str.replace('\%|\%', '', regex=True) ## Remover a porcentagem da variação
cotacao_dolar['Var%'] = pd.to_numeric(cotacao_dolar['Var%'].str.replace(',', '.')) ## Parsear o campo Variação e remover %
cotacao_dolar['Mínima'] = pd.to_numeric(cotacao_dolar['Mínima'].str.replace(',', '.')) ## Parsear o campo Mínima

## Realizar novamente a leitura dos Dados disponibilizados pelo dataset do Case
df_case01 = pd.read_excel('./input/case_analytics-ds.xlsx', sheet_name='Dados')
df_case01.columns = ['Data', 'Canal de Venda', 'Local de Venda', 'PAX', 'Monetário', 'RPK'] ## Renomear as colunas do dataframe
## Realizar a união dos datasets do Case e da cotação do dólar através do campo data
df_correlacao_dolar = pd.concat([df_case01, cotacao_dolar], join='inner', axis=1)
df_correlacao_dolar
```



	Data	Último	Abertura	Máxima	Mínima	Var%
0	2017-04-25	3.1471	3,1283	3.1712	3.1275	0.61
1	2017-04-24	3.1280	3,1506	3.1506	3.1182	-0.64
2	2017-04-21	3.1482	3,1479	3.1488	3.1479	-0.01
3	2017-04-20	3.1486	3,1507	3.1705	3.1324	-0.11
4	2017-04-19	3.1522	3,1075	3.1545	3.1075	1.42
...
121	2016-07-11	3.2020	3,2348	3.2366	3.1885	-0.98
122	2016-04-11	3.2337	3,2423	3.2649	3.2169	-0.27
123	2016-03-11	3.2425	3,2358	3.2565	3.2210	0.20
124	2016-02-11	3.2359	3,2370	3.2425	3.2355	-0.02
125	2016-01-11	3.2367	3,1888	3.2553	3.1780	1.49

Dataset disponibilizado pela br.investing.com



Dataset disponibilizado pelo estudo de caso

	Data	Canal de Venda	Local de Venda	PAX	Monetário	RPK	Data	Último	Abertura	Máxima	Mínima	Var%
0	2016-11-01	Telégrafo	Mandscape	30	2188.109487	3476	2017-04-25	3.1471	3,1283	3.1712	3.1275	0.61
1	2016-11-01	TeleVenda	Arena	52	3302.375951	8197	2017-04-24	3.1280	3,1506	3.1506	3.1182	-0.64
2	2016-11-01	TeleVenda	Mandscape	19	1041.544400	2425	2017-04-21	3.1482	3,1479	3.1488	3.1479	-0.01
3	2016-11-01	Porta a Porta	Mandscape	496	15566.340547	53278	2017-04-20	3.1486	3,1507	3.1705	3.1324	-0.11
4	2016-11-01	TeleVenda	Vast	46	2721.427289	2039	2017-04-19	3.1522	3,1075	3.1545	3.1075	1.42
...
121	2016-11-08	Porta a Porta	Mandscape	1118	31953.297811	113157	2016-07-11	3.2020	3,2348	3.2366	3.1885	-0.98
122	2016-11-08	TeleVenda	Mandscape	45	2433.649660	5716	2016-04-11	3.2337	3,2423	3.2649	3.2169	-0.27
123	2016-11-08	TeleVenda	Ellipsis	935	18316.166113	50592	2016-03-11	3.2425	3,2358	3.2565	3.2210	0.20
124	2016-11-08	TeleVenda	Vast	108	6365.224509	4713	2016-02-11	3.2359	3,2370	3.2425	3.2355	-0.02
125	2016-11-08	Porta a Porta	Vast	1962	30148.736786	82711	2016-01-11	3.2367	3,1888	3.2553	3.1780	1.49

Ao lado, vemos que as únicas correlações consideradas moderadas devido ao valor R > 0.75, são as que estão no mesmo conjunto de dados.

Portanto, não há uma correlação entre o índice do Dólar e a quantidade de passageiros; RPK e monetário.



Figura 7. Correlação entre as variáveis de RPK e ASK. (Fonte: Autor, 2021)



CASE 02

Análise Exploratória do Banco de Dados Estatístico da ANAC

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Quantidade de passageiros transportados pela Azul, TAM, GOL e Passaredo

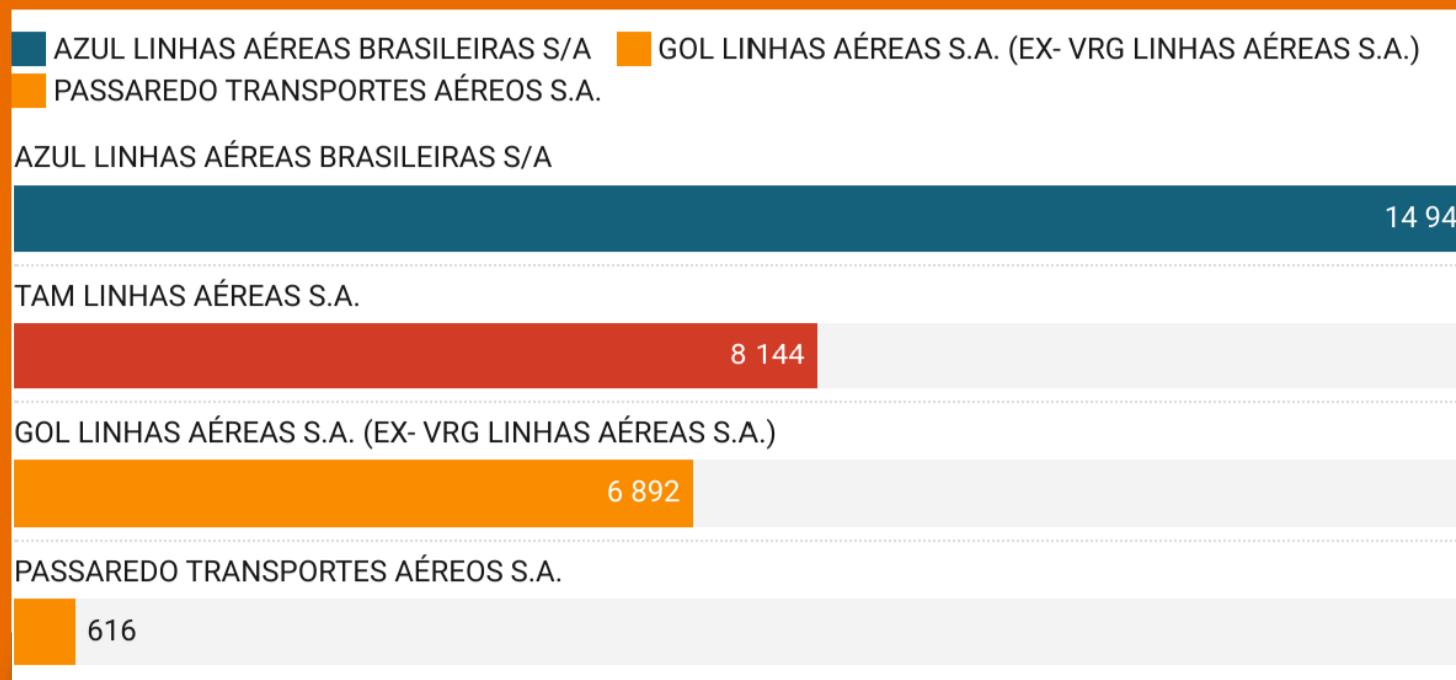


Figura 8. Gráfico de quantidade de passageiros transportados em Maio de 2021. (Fonte: ANAC)

Em Maio, a retomada da aviação foi liderada pela Azul Linhas Aéreas, com mais de 14.943 voos realizados, seguida pela LATAM Linhas Aéreas e em terceiro: a Gol Linhas Aéreas.

Indicadores de RPK e ASK entre Azul, TAM e GOL

Medida	AZUL LINHAS AÉREAS BRASILEIRAS S/A	GOL LINHAS AÉREAS S.A. (EX- VRG LINHAS AÉREAS S.A.)	TAM LINHAS AÉREAS S.A.
Soma de RPK	1 531 705 713	1 219 640 321	1 445 463 066
Média de RPK	102 503	176 964	177 488
Soma de ASK	2 020 426 449	1 385 225 282	2 029 668 313
Média de ASK	135 208	200 990	249 222

Figura 9. Quadro de soma e média de indicadores RPK e ASK registrados em Maio de 2021. (Fonte: ANAC)

Analisando os indicadores que referem a demanda do transporte aéreo estabelecidos pela OACI como RPK - métrica que deve mostrar o número de quilômetros percorridos pelos passageiros pagantes, feito pela multiplicação da quantidade de passageiros pela distância total percorrida e ASK - Assentos disponíveis por quilômetros, apesar da Azul ter um maior número de RPK e ASK tendo em vista que seus vôos internacionais, como por exemplo para o continente americano e Europa, o que fez com que a soma dos valores de RPK e ASK as mantivesse a Azul em primeiro lugar, a os valores pela média acabam reposicionando a Gol Linhas Aéreas em segunda posição.

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

TOP 15 de voos realizados pela Azul Linhas Aéreas

IATA Origem	Cidade de Origem	IATA Destino	Cidade de Destino	Quantidade de Voos
CNF	CONFINS	VCP	CAMPINAS	224
VCP	CAMPINAS	CNF	CONFINS	219
POA	PORTO ALEGRE	VCP	CAMPINAS	216
VCP	CAMPINAS	POA	PORTO ALEGRE	215
VCP	CAMPINAS	SDU	RIO DE JANEIRO	209
SDU	RIO DE JANEIRO	VCP	CAMPINAS	206
SDU	RIO DE JANEIRO	CGH	SÃO PAULO	185
CGH	SÃO PAULO	SDU	RIO DE JANEIRO	183
VCP	CAMPINAS	CWB	SÃO JOSÉ DOS PINHAIS	168
CWB	SÃO JOSÉ DOS PINHAIS	VCP	CAMPINAS	168
SDU	RIO DE JANEIRO	CNF	CONFINS	157
CNF	CONFINS	SDU	RIO DE JANEIRO	155
VCP	CAMPINAS	BSB	BRASÍLIA	150
BSB	BRASÍLIA	VCP	CAMPINAS	148

Figura 10. Quadro de quantidade de trechos realizados pela Azul. (Fonte: ANAC)

TOP 15 de voos realizados pela GOL Linhas Aéreas

IATA Origem	Cidade de Origem	IATA Destino	Cidade de Destino	Quantidade de Voos
SDU	RIO DE JANEIRO	CGH	SÃO PAULO	168
CGH	SÃO PAULO	SDU	RIO DE JANEIRO	166
GRU	GUARULHOS	BSB	BRASÍLIA	102
BSB	BRASÍLIA	GRU	GUARULHOS	100
CNF	CONFINS	GRU	GUARULHOS	98
REC	RECIFE	GRU	GUARULHOS	98
GRU	GUARULHOS	REC	RECIFE	97
POA	PORTO ALEGRE	GRU	GUARULHOS	97
GRU	GUARULHOS	CNF	CONFINS	97
FLN	FLORIANÓPOLIS	GRU	GUARULHOS	96
GRU	GUARULHOS	FLN	FLORIANÓPOLIS	96
GRU	GUARULHOS	POA	PORTO ALEGRE	96
GRU	GUARULHOS	FOR	FORTALEZA	95
FOR	FORTALEZA	GRU	GUARULHOS	95
CWB	SÃO JOSÉ DOS PINHAIS	GRU	GUARULHOS	94

Figura 11. Quadro de quantidade de trechos realizados pela GOL. (Fonte: ANAC)

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

Tanto a Azul Linhas Aéreas quanto a Gol linhas Aéreas, utilizam o paradigma spoke-hub de distribuição de voos. O paradigma consiste em centralizar os voos em Aeroportos onde há maior número de voos saindo para atender outras cidades, então, os aeroportos onde a movimentação não é tão expressiva deve ser ligados em um Aeroporto onde há o maior número de voos. O spoke-hub está mais presente na Azul, uma vez que seu HUB principal é VCP. Enquanto na Gol, o HUB é GRU e CGH. Diferente da Gol, a Azul teve um número mais expressivo em trechos comuns entre o HUB ao Aeroporto de Confins (CNF). Isso é devido a Azul ter um expressivo número de voos regionais interligado o interior de Minas Gerais a capital e também ser um HUB estratégico da empresa. Uma outra análise é do trecho mais comum do Brasil: a ponte aérea. A Azul Linhas Aéreas realizou mais de 185 voos entre Santos Dumont (RJ) para Congonhas (SP), enquanto a Gol realizou 168.

Em Maio de 2021, a GOL Linhas Aéreas registrou 168 voos entre Santos Dumont (SDU) para Congonhas (CGH). Entretanto, não foi o maior operador. A Azul Linhas Aéreas registrou 185 voos.



Analise exploratória nos voos da Ponte Aérea

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Embora tenha mais voos, a GOL tem a maior ocupação média (considerando RPK/ASK), atingindo cerca de 83.19%, contra 70.57% da Azul Linhas Aéreas.

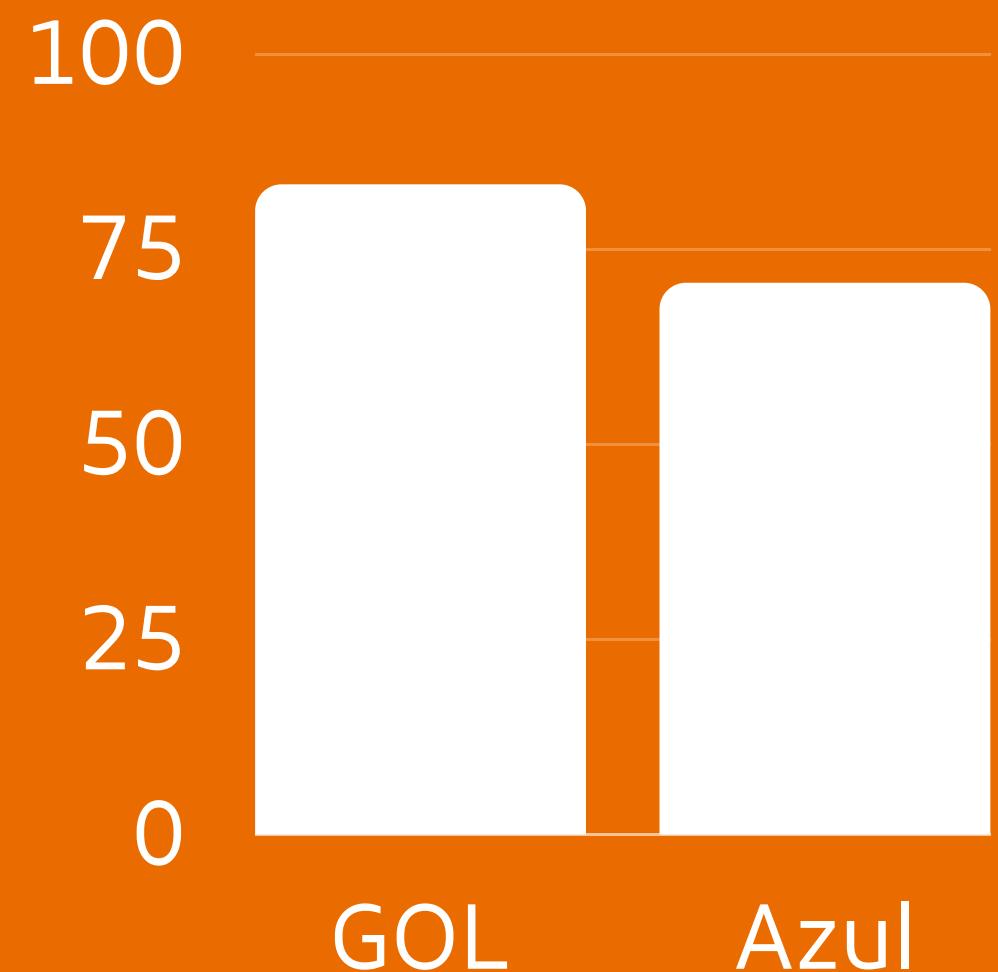


Figura 13. Gráfico de média de ocupação nos trechos de Santos Dumont para Congonhas registrados pela GOL e Azul em Maio de 2021. (Fonte: ANAC)

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Embora tenha mais voos, a GOL tem a maior ocupação média (considerando RPK/ASK), atingindo cerca de 83.19%, contra 70.57% da Azul Linhas Aéreas.

	AZUL Linhas Aéreas	GOL Linhas Aéreas
count	185.000000	166.000000
mean	70.573135	83.191265
std	19.433047	11.131899
min	33.050000	53.450000
25%	52.540000	76.605000
50%	72.030000	85.385000
75%	88.980000	92.385000
max	100.000000	99.440000

Figura 14. Quadro da estatística descritiva em relação a ocupação dos voos realizados entre o trecho SDU (Santos Dumont) para CGH (Congonhas). (Fonte: ANAC)

Processo Seletivo Gol Linhas Aéreas

Desenvolvedor (E)

nm_empresa	AZUL LINHAS AÉREAS BRASILEIRAS S/A							GOL LINHAS AÉREAS S.A. (EX- VRG LINHAS AÉREAS S.A.)							
	diasemana_partida_real	Domingo	Quarta-feira	Quinta-feira	Segunda-feira	Sexta-feira	Sábado	Terça-feira	Domingo	Quarta-feira	Quinta-feira	Segunda-feira	Sexta-feira	Sábado	Terça-feira
count		28.000000	22.000000	28.000000	35.000000	30.000000	19.000000	23.000000	26.000000	19.000000	21.000000	33.000000	28.000000	21.000000	18.000000
mean		89.589286	63.518182	55.503571	82.825714	65.226667	61.057895	68.682609	89.234615	77.078947	83.642857	86.133333	80.110714	79.657143	83.955556
std		7.194151	20.006895	13.450223	13.465388	17.478005	19.815687	18.098283	8.939483	12.089875	9.503345	11.998090	12.122594	10.351598	6.858762
min		69.500000	33.100000	35.600000	52.500000	34.700000	34.700000	41.500000	61.600000	55.200000	58.200000	53.400000	53.800000	60.100000	69.000000
25%		86.875000	45.325000	44.900000	70.300000	47.500000	44.100000	54.200000	86.750000	67.100000	79.700000	82.300000	73.525000	71.300000	80.275000
50%		91.500000	63.150000	52.500000	89.000000	67.350000	57.600000	72.000000	91.950000	80.500000	85.300000	90.600000	81.500000	83.700000	84.600000
75%		94.300000	80.750000	59.550000	91.950000	78.600000	76.250000	83.900000	94.475000	85.750000	91.300000	95.300000	89.925000	87.600000	88.950000
max		100.000000	96.600000	92.400000	97.500000	94.900000	93.200000	97.500000	98.900000	96.600000	96.700000	99.400000	95.400000	93.000000	95.000000

Figura 15. Estatística descritiva para a taxa de ocupação dos voos realizados entre Santos Dumont e Congonhas pela Azul Linhas Aéreas e GOL Linhas Aéreas por dia da semana. (Fonte: ANAC)

Em uma análise a média de ocupação nos voos entre SDU e CGH é possível identificar que Domingo e Segunda são os dias mais críticos para a operação da ponte aérea para as duas empresas, sendo que a Gol tem obtido maior ocupação na Segunda-Feira, com 86.14, ao contrário da Azul, que teve 82.83%.

Em Maio de 2021, a GOL registrou 19 destinos com a média de ocupação menor de 50%. Embora alguns destinos sejam consideráveis como Parauapebas (CKS), onde foi o mês de estreia da companhia e outros onde há medidas sanitárias que dificultam a melhoria nos números de ocupação.



Destinos GOL com baixa taxa de ocupação

Ocupação dos destinos GOL em Maio de 2021.

sg_iata_origem	nm_municipio_origem	sg_iata_destino	nm_municipio_destino	ocupacao
FOR	FORTALEZA	BVB	BOA VISTA	0.000000
BAQ	BARRANQUILLA	GYE	GUAYAQUIL	0.500000
CKS	PARAUAPEBAS	SLZ	SÃO LUÍS	27.150000
SLZ	SÃO LUÍS	CKS	PARAUAPEBAS	28.200000
GIG	RIO DE JANEIRO	EZE	EZEIZA, BUENOS AIRES PROVINCE	29.000000
GYE	GUAYAQUIL	GIG	RIO DE JANEIRO	29.000000
GIG	RIO DE JANEIRO	BAQ	BARRANQUILLA	29.600000
EZE	EZEIZA, BUENOS AIRES PROVINCE	GIG	RIO DE JANEIRO	29.600000
GIG	RIO DE JANEIRO	SCL	SANTIAGO	30.600000
SCL	SANTIAGO	GIG	RIO DE JANEIRO	30.600000
ASU	ASUNCIÓN	CNF	CONFINS	31.700000
CNF	CONFINS	ASU	ASUNCIÓN	31.700000
RBR	RIO BRANCO	CZS	CRUZEIRO DO SUL	40.900000
VCP	CAMPINAS	SDU	RIO DE JANEIRO	43.960000
CNF	CONFINS	CKS	PARAUAPEBAS	44.166667
CZS	CRUZEIRO DO SUL	RBR	RIO BRANCO	44.311765
SDU	RIO DE JANEIRO	VCP	CAMPINAS	45.000000
CKS	PARAUAPEBAS	CNF	CONFINS	45.850000

Figura 16. Voos da GOL Linhas Aéreas que possuem média de ocupação menor que 50%. (Fonte: ANAC)



**Destinos GOL com baixa taxa
de ocupação**

Em Maio de 2021, a GOL registrou 19 destinos com a média de ocupação menor de 50%. Embora alguns destinos sejam consideráveis como Parauapebas (CKS), onde foi o mês de estreia da companhia e outros onde há medidas sanitárias que dificultam a melhoria nos números de ocupação.



Destinos GOL com baixa taxa de ocupação

Ocupação dos destinos GOL em Maio de 2021.

Em Maio de 2021, a GOL registrou 81 destinos com a média de ocupação maior que 90%. Sendo os trechos destaques: SDU/CWB, NAT/CNF, MAO/RBR, JDO/GRU e GRU/JDO.



Destinos GOL com alta taxa de ocupação

Ocupação dos destinos GOL em Maio de 2021.

sg_iata_origem	nm_municipio_origem	sg_iata_destino	nm_municipio_destino	ocupacao
SDU	RIO DE JANEIRO	CWB	SÃO JOSÉ DOS PINHAIS	98.180000
NAT	SÃO GONÇALO DO AMARANTE	CNF	CONFINS	97.300000
MAO	MANAUS	RBR	RIO BRANCO	97.300000
JDO	JUAZEIRO DO NORTE	GRU	GUARULHOS	96.685714
GRU	GUARULHOS	JDO	JUAZEIRO DO NORTE	96.607407
NVT	NAVEGANTES	CNF	CONFINS	96.200000
SDU	RIO DE JANEIRO	POA	PORTO ALEGRE	96.169444
IOS	ILHÉUS	GRU	GUARULHOS	96.021429
PVH	PORTO VELHO	BSB	BRASÍLIA	95.937500
BSB	BRASÍLIA	CGR	CAMPO GRANDE	95.900000
		REC	RECIFE	95.772222
SDU	RIO DE JANEIRO	SSA	SALVADOR	95.764516
CGB	VÁRZEA GRANDE	BSB	BRASÍLIA	95.709524
REC	RECIFE	GRU	GUARULHOS	95.358163
CGB	VÁRZEA GRANDE	GRU	GUARULHOS	95.208824
FLN	FLORIANÓPOLIS	POA	PORTO ALEGRE	95.200000
BPS	PORTO SEGURO	CNF	CONFINS	95.200000
GRU	GUARULHOS	CGB	VÁRZEA GRANDE	95.076471
		POA	PORTO ALEGRE	94.878125
SDU	RIO DE JANEIRO	FLN	FLORIANÓPOLIS	94.700000
CNF	CONFINS	AJU	ARACAJU	94.600000
REC	RECIFE	BSB	BRASÍLIA	94.544444
BSB	BRASÍLIA	POA	PORTO ALEGRE	94.541667

Figura 17. Voos da GOL Linhas Aéreas que possuem média de ocupação maior que 90%. (Fonte: ANAC)



**Destinos GOL com alta taxa de
ocupação**

Ocupação dos destinos GOL em Maio de 2021.

sg_iata_origem	nm_municipio_origem	sg_iata_destino	nm_municipio_destino	ocupacao
SDU	RIO DE JANEIRO	CWB	SÃO JOSÉ DOS PINHAIS	98.180000
NAT	SÃO GONÇALO DO AMARANTE	CNF	CONFINS	97.300000
MAO	MANAUS	RBR	RIO BRANCO	97.300000
JDO	JUAZEIRO DO NORTE	GRU	GUARULHOS	96.685714
GRU	GUARULHOS	JDO	JUAZEIRO DO NORTE	96.607407
NVT	NAVEGANTES	CNF	CONFINS	96.200000
SDU	RIO DE JANEIRO	POA	PORTO ALEGRE	96.169444
IOS	ILHÉUS	GRU	GUARULHOS	96.021429
PVH	PORTO VELHO	BSB	BRASÍLIA	95.937500
BSB	BRASÍLIA	CGR	CAMPO GRANDE	95.900000
		REC	RECIFE	95.772222
SDU	RIO DE JANEIRO	SSA	SALVADOR	95.764516
CGB	VÁRZEA GRANDE	BSB	BRASÍLIA	95.709524
REC	RECIFE	GRU	GUARULHOS	95.358163
CGB	VÁRZEA GRANDE	GRU	GUARULHOS	95.208824
FLN	FLORIANÓPOLIS	POA	PORTO ALEGRE	95.200000
BPS	PORTO SEGURO	CNF	CONFINS	95.200000
GRU	GUARULHOS	CGB	VÁRZEA GRANDE	95.076471
		POA	PORTO ALEGRE	94.878125
SDU	RIO DE JANEIRO	FLN	FLORIANÓPOLIS	94.700000
CNF	CONFINS	AJU	ARACAJU	94.600000
REC	RECIFE	BSB	BRASÍLIA	94.544444
BSB	BRASÍLIA	POA	PORTO ALEGRE	94.541667

Figura 18. Voos da GOL Linhas Aéreas que possuem média de ocupação maior que 90%. (Fonte: ANAC)



**Destinos GOL com alta taxa de
ocupação**

Para a análise, foi necessário encontrar qual o trecho a Gol mais operou em Maio de 2021.

Apesar de encontrar os trechos CGH/SDU e SDU/CGH, resolvi escolher o trecho GRU-BSB que teve cerca de 100 voos realizados.



Correlação faixa de Horário x
indicador de RPK e ASK

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

O trecho GRU-BSB teve cerca de 100 voos, sendo Segunda-Feira e Domingo onde teve a maior quantidade de voos, respectivamente sendo 18 e 16.

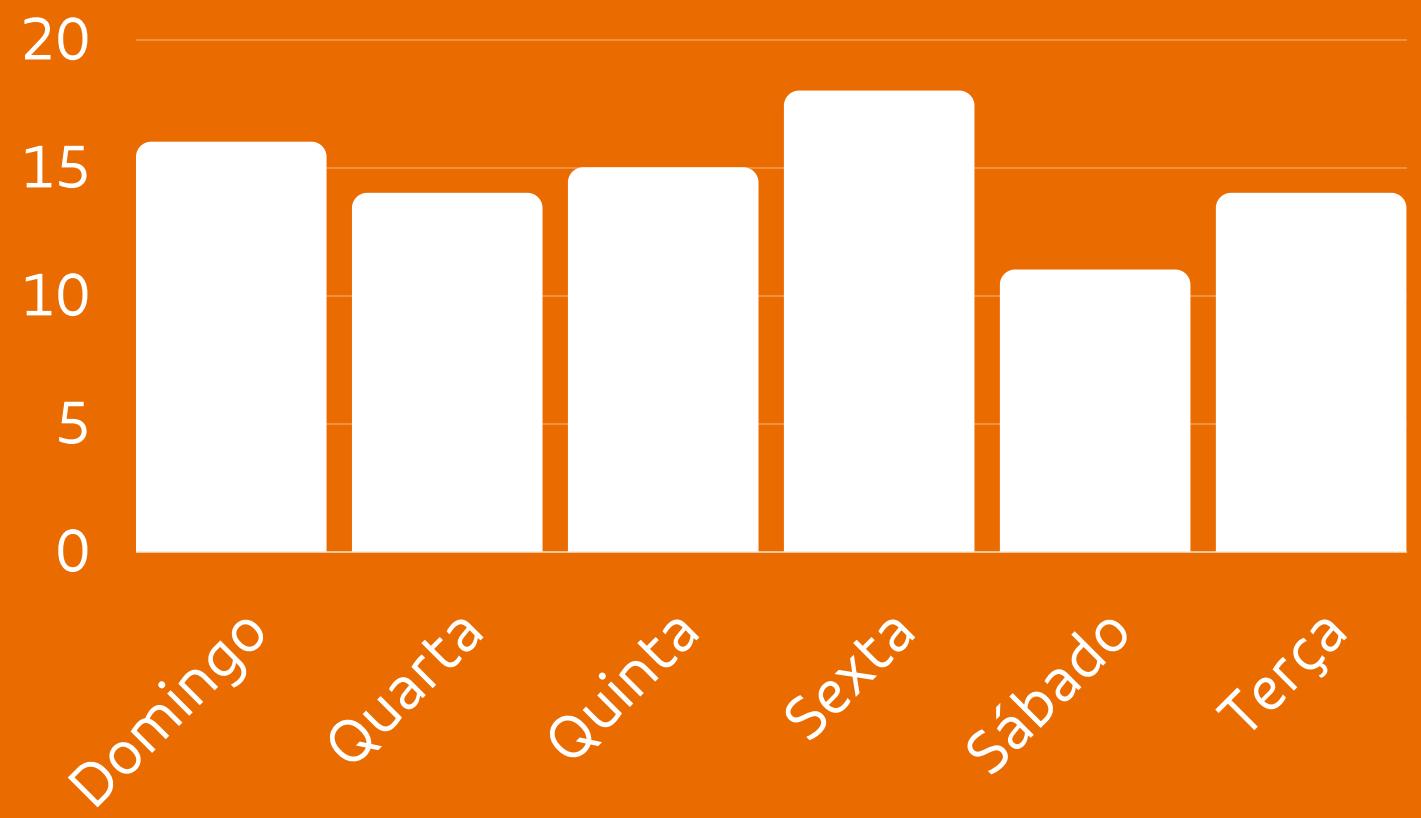


Figura 19. Distribuição da quantidade de voos realizados por dia da semana entre o trecho Guarulhos para Brasília. (Fonte: ANAC)

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Após isso, analisei qual foi a distribuição de quantidade de voos pela faixa de horários nos voos que são realizados entre GRU-BSB na Segunda-Feira.

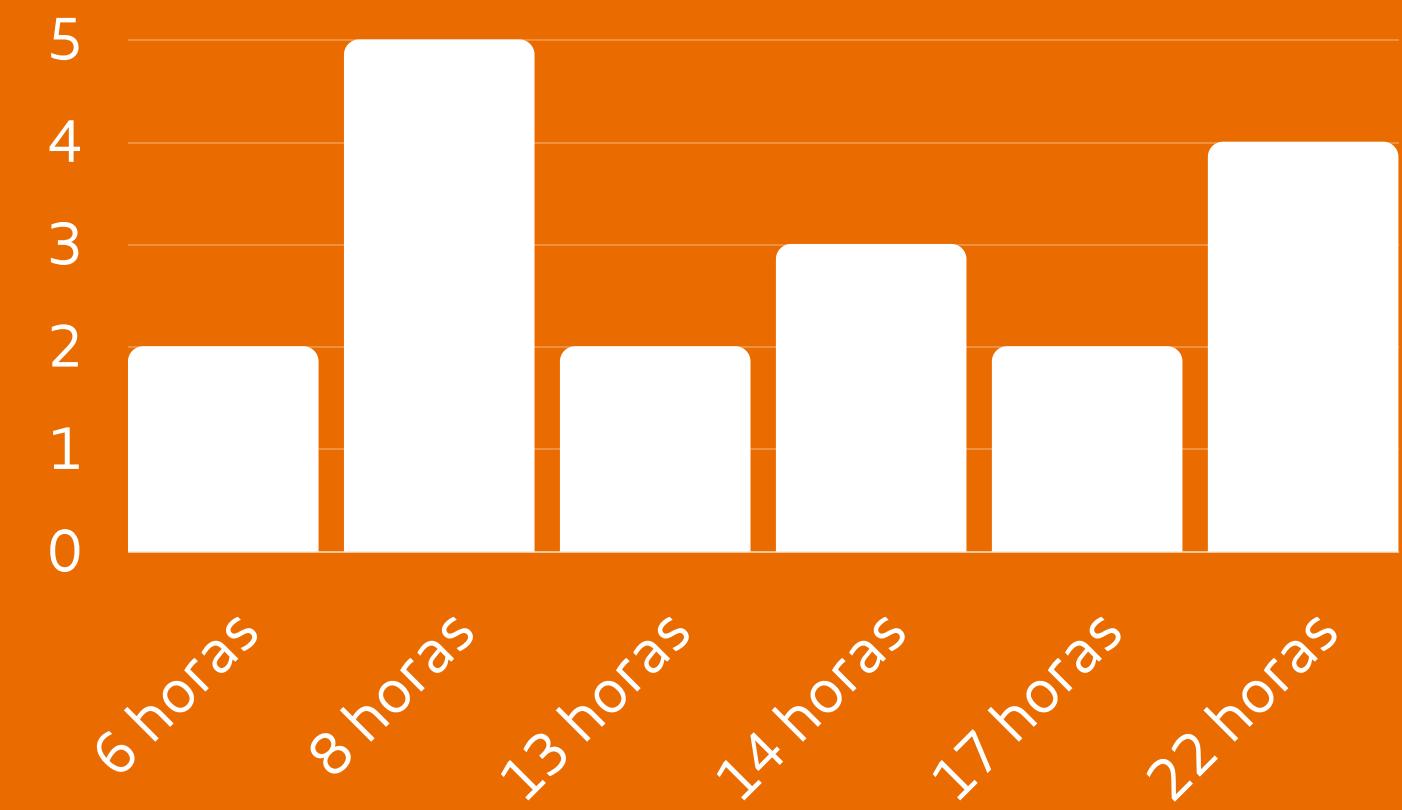


Figura 20. Distribuição da quantidade de voos realizados por faixa de horário, entre o trecho Guarulhos para Brasília, na segunda-feira. (Fonte: ANAC)

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Em sequência, analisei os valores referentes a soma de RPK e ASK na segunda-feira pela faixa de horário.

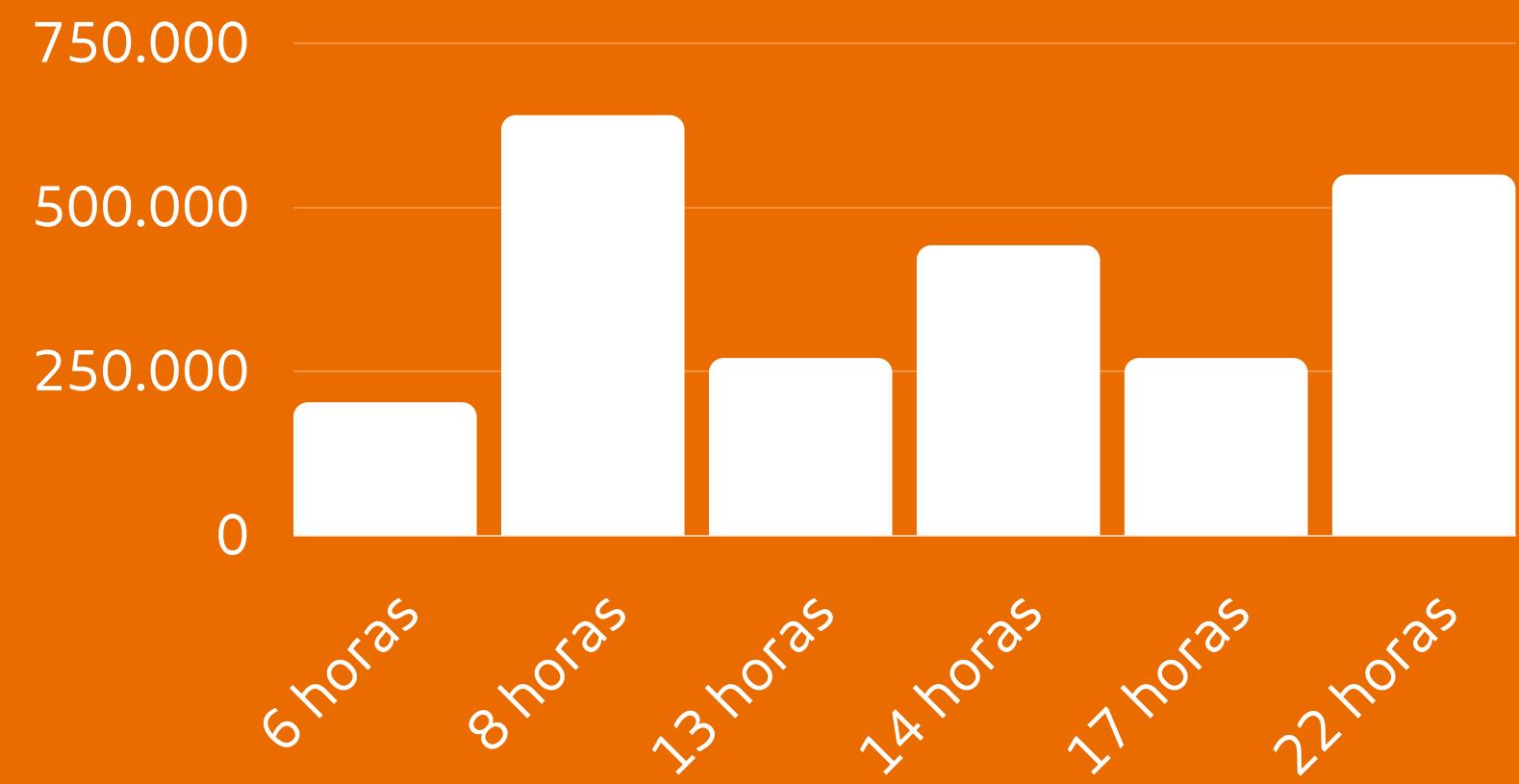


Figura 21. Distribuição da soma de RPK por faixa de horário, entre o trecho Guarulhos para Brasília, na segunda-feira. (Fonte: ANAC)

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

O maior pico está entre 8 horas e 22 horas, sendo que às 08 na segunda-feira, a GOL registrou 639K, enquanto às 22h registrou 548K.

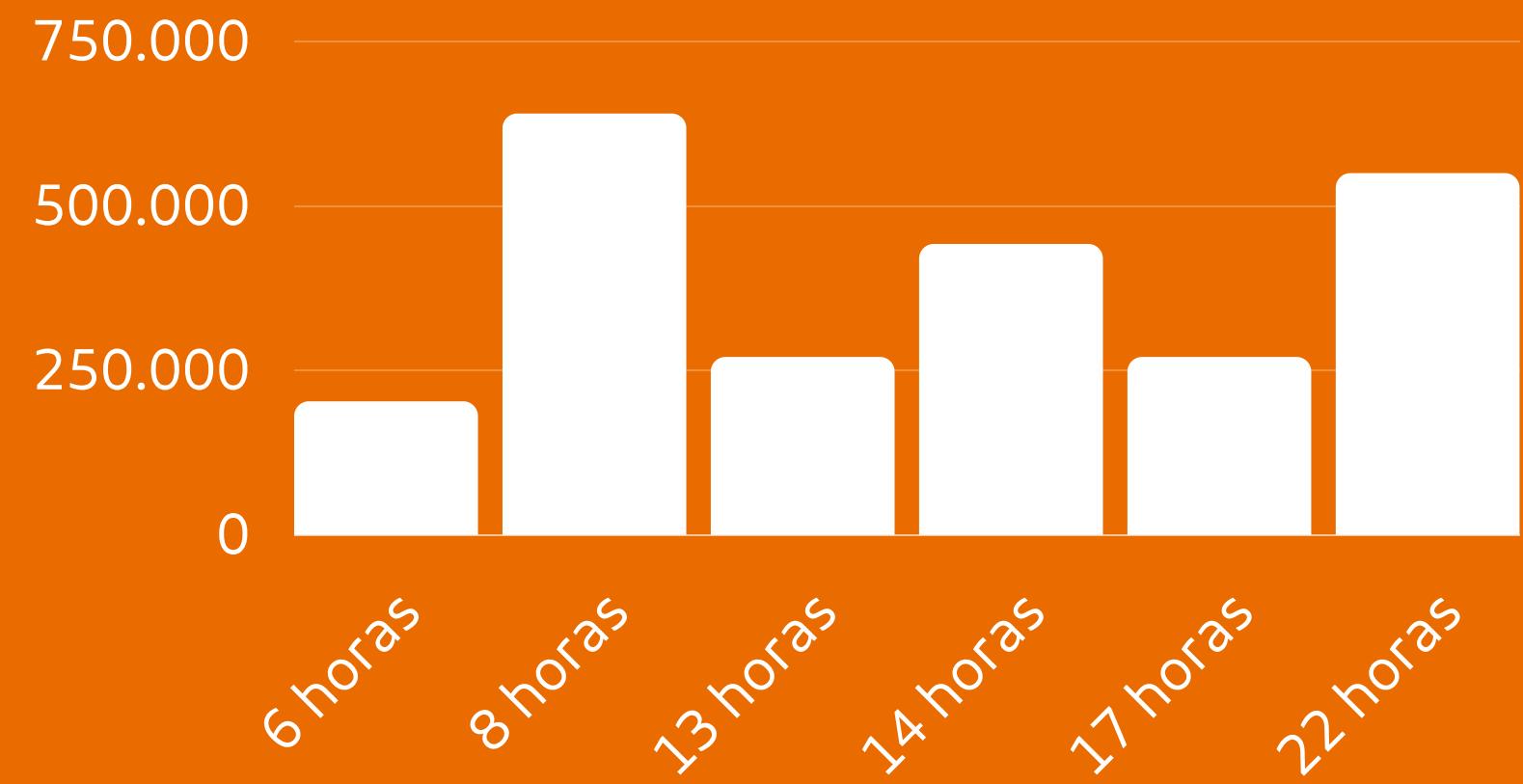


Figura 22. Distribuição da soma de RPK por faixa de horário, entre o trecho Guarulhos para Brasília, na segunda-feira. (Fonte: ANAC)

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Em sequência, realizei a correlação linear, usando correlação de sp entre o horário de Spearman, uma vez que a distribuição é não normal, entre a partida de cada voo, soma do RPK e soma do ASK.

	hr_partida_real	nr_rpk	nr_ask
hr_partida_real	1.000000	0.205703	0.045707
nr_rpk	0.205703	1.000000	0.981993
nr_ask	0.045707	0.981993	1.000000

Figura 23. Correlação entre a faixa de horário, soma de RPK e ASK, entre o trecho Guarulhos-Brasília, realizados na segunda-feira. (Fonte: A)

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Através do resultado obtido, a correlação de pearson entre a soma de RPK e a faixa de horário existe, entretanto, é fraca. Já a correlação entre horário de partida e ASK é desprezível

	hr_partida_real	nr_rpk	nr_ask
hr_partida_real	1.000000	0.205703	0.045707
nr_rpk	0.205703	1.000000	0.981993
nr_ask	0.045707	0.981993	1.000000

Figura 24. Correlação entre a faixa de horário, soma de RPK e ASK, entre o trecho Guarulhos-Brasília, realizados na segunda-feira. (Fonte: ANAC)

Valores expressos em p-valor e r-valor. Análise de correlação de Speraman. Para a classificação do R, utilizou-se a referência de Vieira(2011). Teste de normalidade Shapiro-wilk teste p>0.43 para nr_rpk e 0.281 para nr_ask.

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Entretanto, se fizermos uma correlação entre a média de RPK e ASK com a faixa de horário, encontraremos uma correlação forte ($p = 0.782$).

	hr_partida_real	nr_rpk	nr_ask
hr_partida_real	1.000000	0.782691	0.169031
nr_rpk	0.782691	1.000000	0.342997
nr_ask	0.169031	0.342997	1.000000

Figura 25. Correlação entre a faixa de horário, média de RPK e ASK, entre o trecho Guarulhos-Brasília, realizados na segunda-feira. (Fonte: ANAC)

Valores expressos em p-valor e r-valor. Análise de correlação de Speraman. Para a classificação do R, utilizou-se a referência de Vieira(2011). Teste de normalidade Shapiro-wilk teste $p>0.13$ para nr_rpk e 0.006 para nr_ask.

Processo Seletivo Gol Linhas Aéreas Desenvolvedor (E)

Portanto, podemos considerar que voos noturnos, isso é (18>), maior número de RPK e ASK podemos obter. Isso é devido ao retorno de pessoas que vão a trabalho para Brasília e devido a um valor menor de tarifa devido ao voo ser noturno.

	hr_partida_real	nr_rpk	nr_ask
hr_partida_real	1.000000	0.682308	0.253132
nr_rpk	0.682308	1.000000	0.797609
nr_ask	0.253132	0.797609	1.000000

Figura 26. Correlação entre a faixa de horário, média de RPK e ASK, entre o trecho Guarulhos-Brasília, realizados na segunda-f



Obrigado!

Conheça a minha pré entrevista em <https://youtu.be/nKq7ylhXQ3c>

Case Analytics

Maycon Felipe Mota – Desenvolvedor (E)

