

Turma 1TSCPG 2025

Equipe 2FT

Fábio Fadul Cunha Yoshida
RM561530

Felipe Andueza Paullelli-
RM559449

Tiago Abou Hamia França -
RM562349



Rota Travelers
Brasil

O desafio de prever
a data e o trecho
da próxima compra na
empresa que mais vende
passagens online no Brasil



O que é necessário para uma boa previsão?

Segmentação realista de
perfis de comportamento;
Tratamento estatístico
adequado.



Para prever bem
Há de segmentar os
perfis de
comportamento
baseada nos dados.



Tratamento inicial
Hash em números
sequenciais.

Segmentação inicial:

~51% compra única

~49% compra múltipla



Solução sugerida de segmentação

Compra única:

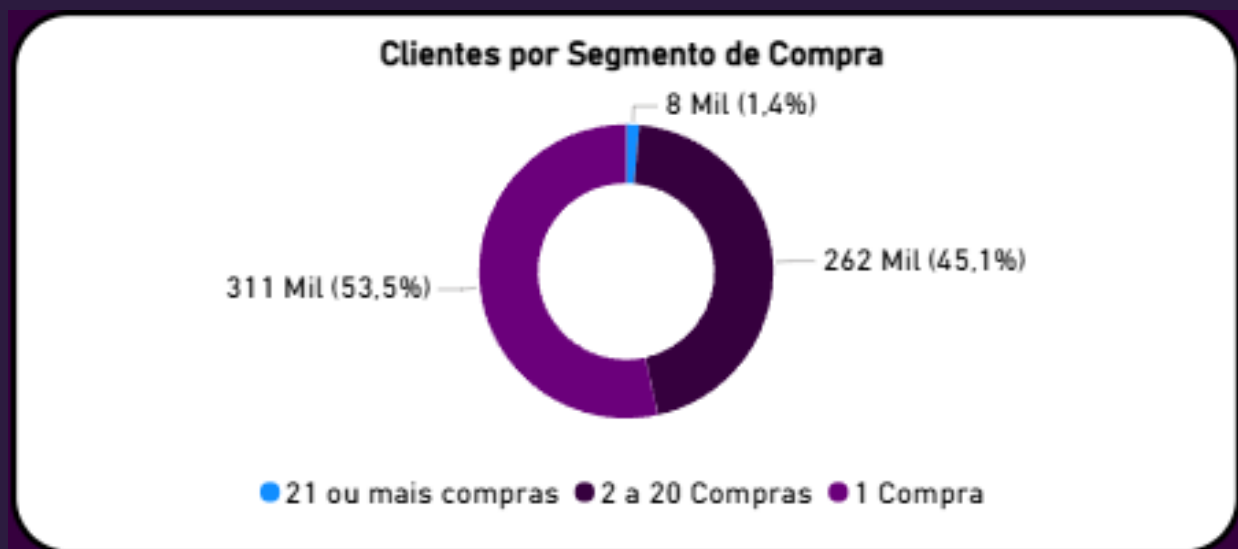
53%

de 2 a 20

compras: 45%

mais de 20

compras: 2%



E há de escolher os
testes estatísticos
adequados de
previsão para cada
caso (trecho e data)



Solução Estatística

Sistema de Recomendação para trecho; Árvore de decisão para data



Solução Estatística

```
# Ordena o dataset por cliente e data/hora da compra
dataset_de2a20_compras = dataset_de2a20_compras.sort_values(by=['id_cliente', 'data_compra(a/m/d)', 'hora_compra(h/m/s)'])

# Pega sequência de destinos de ida para cada cliente
sequences = dataset_de2a20_compras.groupby('id_cliente')['id_local_destino_ida'].apply(list)

# Cria pares de transições (destino_atual -> destino_seguinte) para todos clientes
pairs = []
for seq in sequences:
    for i in range(len(seq)-1):
        pairs.append((seq[i], seq[i+1]))

pairs_df = pd.DataFrame(pairs, columns=['atual', 'seguinte'])

# Cria matriz de transição: para cada destino_atual, quais os destinos_seguinte e frequência
transitions = pairs_df.groupby(['atual', 'seguinte']).size().reset_index(name='freq')

# Normaliza para obter probabilidades de transição
transitions['prob'] = transitions.groupby('atual')['freq'].transform(lambda x: x / x.sum())

# Agora a função que, dado um destino atual, retorna o destino seguinte mais provável
def recomendar_proximo_destino(destino_atual):
    candidatos = transitions[transitions['atual'] == destino_atual]
    if candidatos.empty:
        # Se não houver histórico, recomenda o destino mais frequente geral
        return dataset_de2a20_compras['id_local_destino_ida'].mode()[0]
    else:
        return candidatos.sort_values('prob', ascending=False).iloc[0]['seguinte']
```

Sistema de Recomendação para trecho



Solução Estatística

Árvore de decisão para data

```
# 9. Dividir treino e teste
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 10. Treinar modelo
model_reg = RandomForestRegressor()
model_reg.fit(X_train, y_train)

# 11. Avaliar desempenho
y_pred = model_reg.predict(X_test)
mae = mean_absolute_error(y_test, y_pred)
print(f"Erro médio (MAE): {mae:.2f} dias")
```



Dashboard
Quatro áreas:
Próxima Compra
Próximo Trecho
Escolha de Segmentos
pré-definidos
Quadro de totais



Protótipos da Solução Proposta



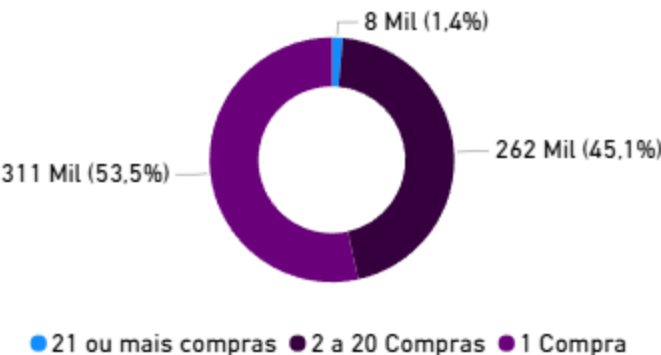
Rota Travelers Brasil

R\$ 161,10
Valor médio do ticket

14:53
Horário de pico de compras

581.817
Clientes Únicos

Clientes por Segmento de Compra



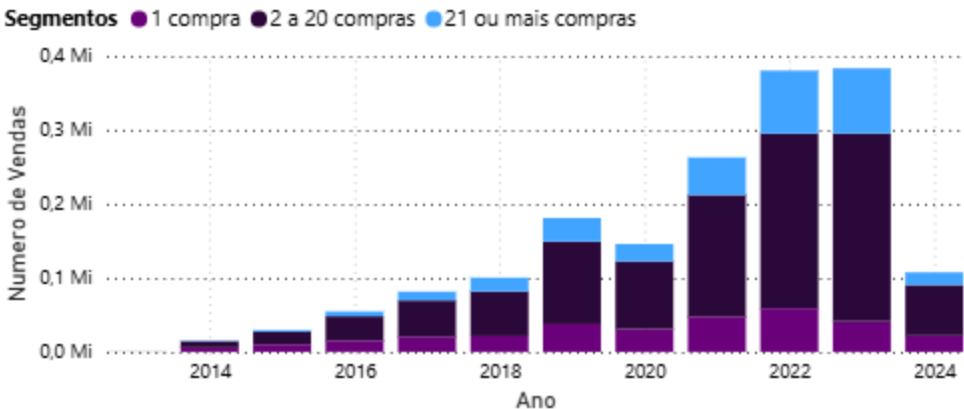
Top 5 destinos por Segmento

São Paulo
Rio de Janeiro
Belo Horizonte
Curitiba
Porto Alegre

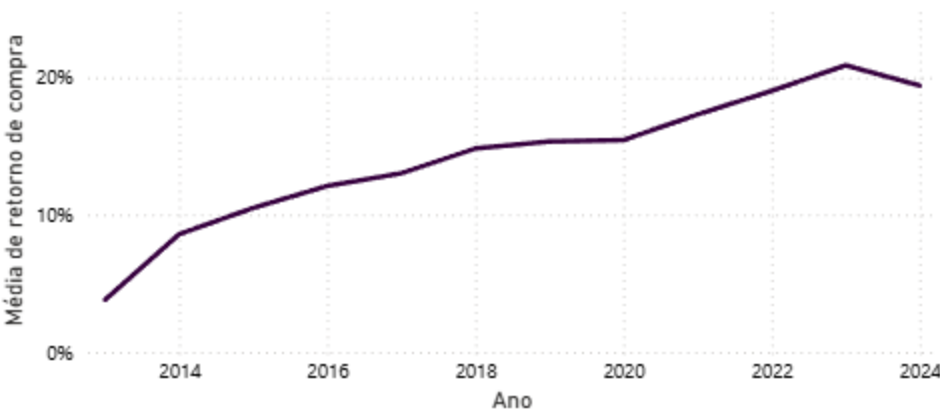
17,86%
Clientes que retornam em 7 a 30 dias

17,54%
Probabilidade geral de retorno

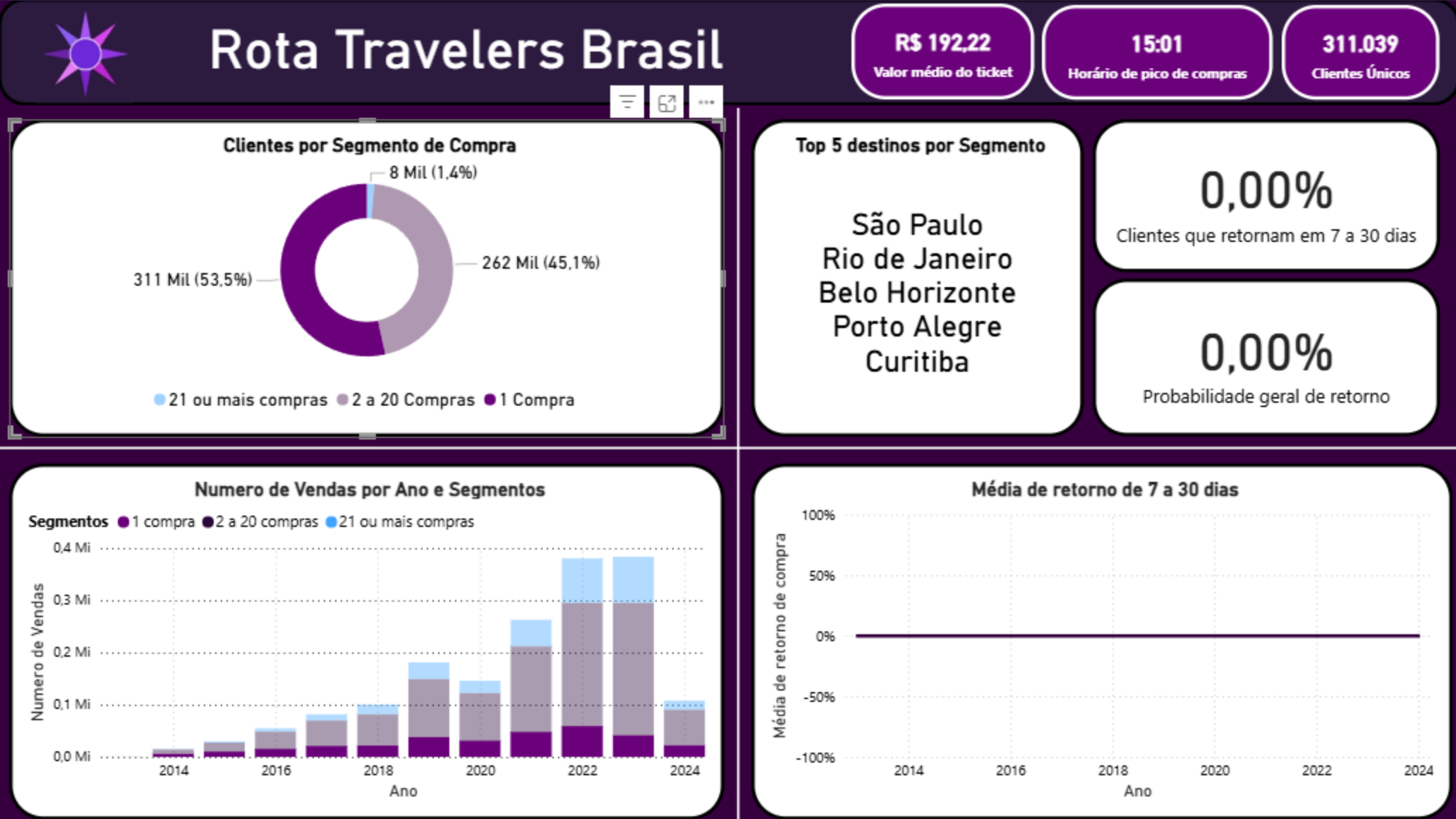
Numero de Vendas por Ano e Segmentos



Média de retorno de 7 a 30 dias



Cientes que compram uma só vez



Clientes que fizeram de 2 a 20 compras



Rota Travelers Brasil

R\$ 161,21

Valor médio do ticket

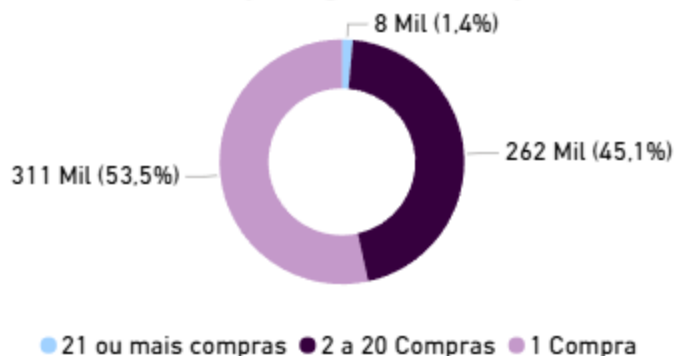
14:56

Horário de pico de compras

262.400

Clientes Únicos

Clientes por Segmento de Compra



Top 5 destinos por Segmento

São Paulo
Rio de Janeiro
Belo Horizonte
Porto Alegre
Curitiba

36,41%

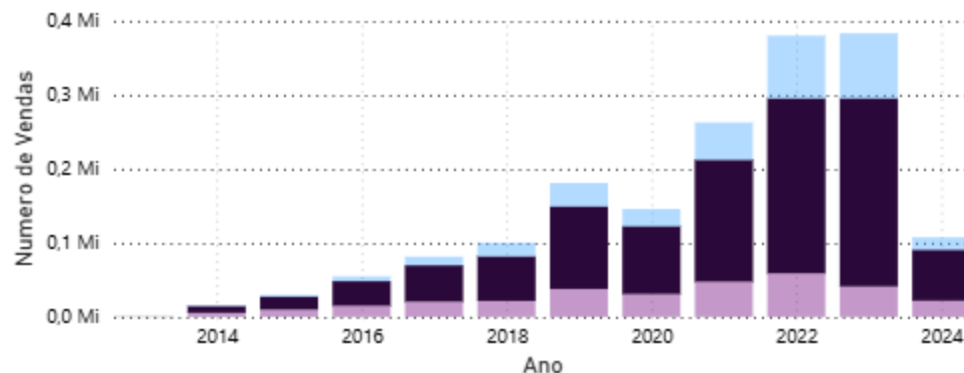
Clientes que retornam em 7 a 30 dias

16,88%

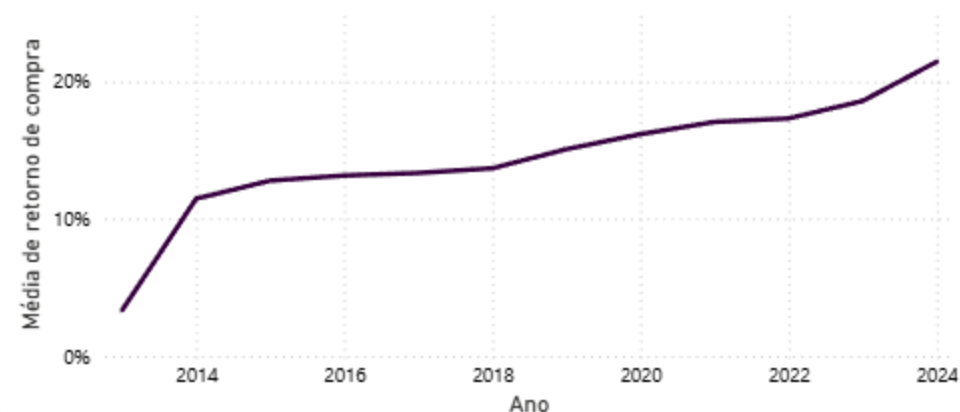
Probabilidade geral de retorno

Numero de Vendas por Ano e Segmentos

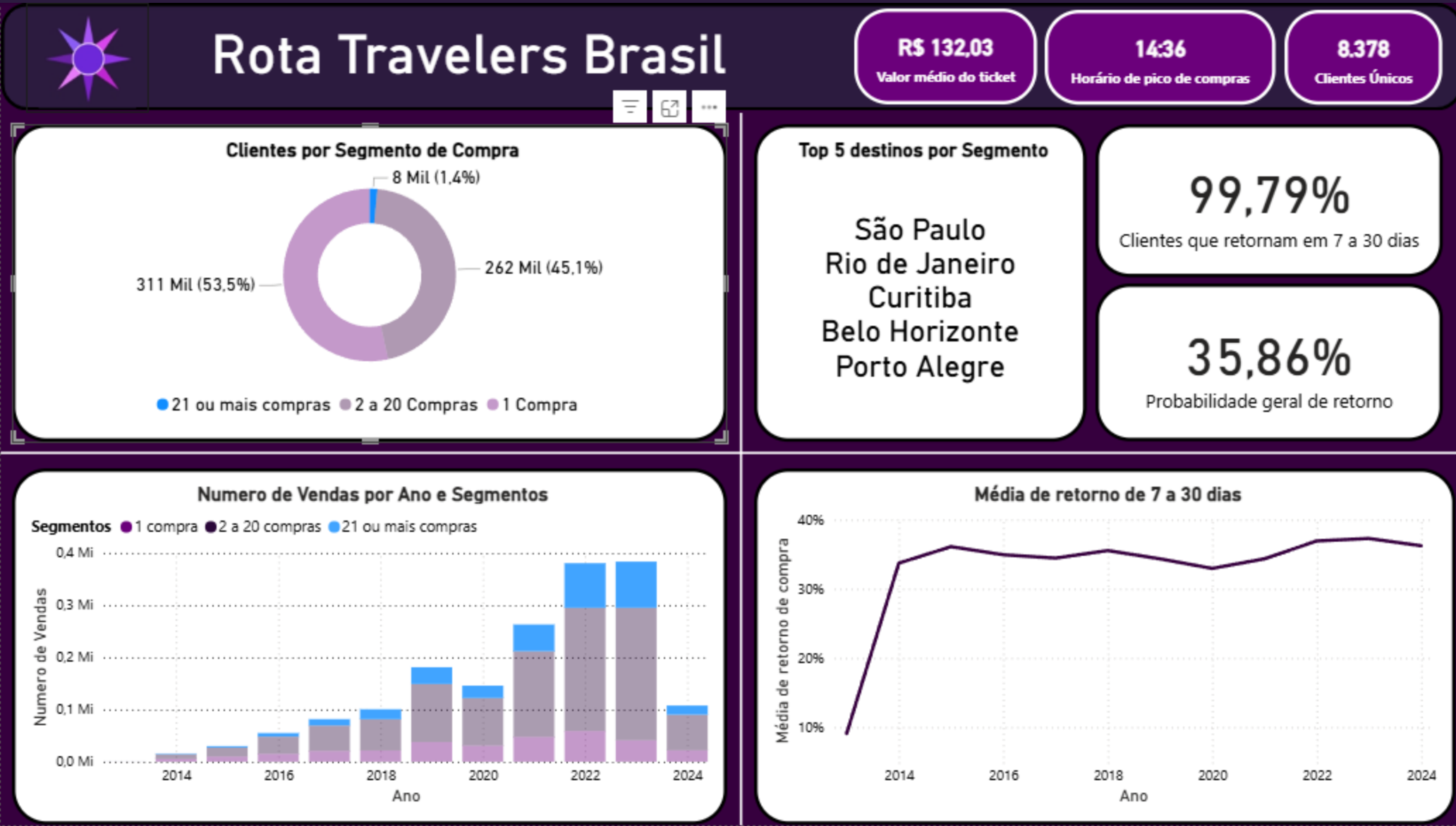
Segmentos ● 1 compra ● 2 a 20 compras ● 21 ou mais compras



Média de retorno de 7 a 30 dias



Clients que fizeram mais de 20 compras



Rota Travelers Brasil Dashboard

https://app.powerbi.com/links/syqAs7ngVI?ctid=11dbbfe2-89b8-4549-be10-cec364e59551&pbi_source=linkShare



Compras volumosas: Programa fidelidade à escolha da Clickbus



Fontes

Estatística e Ciências de Dados - Morettini e Singer

Agência Nacional de Transporte Terrestres

Knime

Manus IA

Jupyter Notebook (Anaconda e Google Colab)

<https://www.datacamp.com/tutorial/random-forests-classifier-python>



Rota Travelers Brasil Pitch

<https://youtu.be/NPktMbPpo7Q>



Agradecimentos

ClickBus – Equipe de Dados

Fiap – Professores do curso de Data Science

Colegas

