



CLUB BUS





→ DataWay

ITSCPv



Enos Barros

RM561926

Isabella Heder

RM561300

João Dalessio

RM561050

Milena Barbosa

RM558913

Paulo Luchini

RM561477

REPRESENTANTE

Nossa Solução Resumida



O que é?

O **Club Bus** é um **programa de fidelidade inteligente gratuito** da ClickBus voltado para entender melhor o perfil dos consumidores e melhorar a experiência dos clientes.

Todas as pessoas que tenham comprado passagens pelo menos uma vez, já fazem parte do Club Bus!

Apresentação do Cenário

○ **Maior Plataforma de Passagens Rodoviárias do Brasil**

- A ClickBus é líder em vendas online de passagens rodoviárias, com **milhões de clientes e dados** valiosos.

Busca:

- Aumentar a recorrência de compras
- Aumentar a fidelização dos clientes
- Conhecer melhor os seus clientes

Falta de conhecimento profundo **sobre o perfil dos clientes** e baixa previsibilidade de suas próximas ações de compra.

Enfrenta **dificuldades em entender o comportamento dos compradores** e ativá-los de forma personalizada.

Principais Problemas

Falta de conhecimento sobre o perfil dos compradores

- Dificulta a criação de estratégias eficazes e impede a personalização da comunicação com o cliente.

Dificuldade em personalizar ofertas e comunicações

- Resulta em campanhas genéricas, com baixo engajamento e menor taxa de conversão.

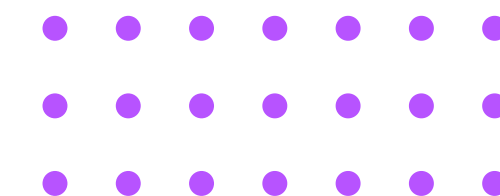
Perda de oportunidades de venda por falta de segmentação

- A empresa deixa de aproveitar momentos estratégicos e perfis específicos que poderiam gerar mais receita.

Impossibilidade de prever comportamentos futuros

- Compromete o planejamento comercial e limita a capacidade de antecipar demandas e necessidades dos clientes.

Objetivo do Projeto



Aumento das Vendas

- Com campanhas mais eficazes e personalizadas, a taxa de conversão e receita tendem a crescer.

Melhor Eficiência em Ações de Marketing

- Redução de desperdício com ações genéricas e foco em estratégias baseadas em dados, convertendo mais clientes em vendas.

Fidelização de Clientes

- Ao se sentirem compreendidos e bem atendidos, os clientes tendem a voltar a comprar com a ClickBus.

Melhoria na Experiência do Cliente

- Ofertas e comunicações mais relevantes tornam a jornada de compra mais satisfatória.

Definindo nossa solução:



O **Club Bus** é um **programa de fidelidade** gratuito da ClickBus, criado para aprofundar o conhecimento sobre os clientes que já realizaram compras na plataforma.

Ele permite **personalizar a experiência dos consumidores** e aumentar as vendas da empresa através do **convertimento do interesse dos clientes, em compras**.

A nossa solução **utiliza dados** de comportamento **para segmentar** os consumidores **e oferecer uma experiência mais personalizada**.

Como vamos alcançar os objetivos?

Mapear e compreender o perfil dos clientes

Utilizamos a base de dados da Click Bus, já tratada, para **identificar padrões** de comportamento e **segmentar clientes em grupos** estratégicos.

Transformar interesse em compras efetivas

A personalização aumenta as chances de conversão, transformando o interesse inicial em vendas reais, ao **indicar destinos e rotas de maior afinidade**.

Oferecer recomendações e ofertas personalizadas

Com a análise dos dados, conseguimos criar **sugestões relevantes no site** adaptando as ofertas ao histórico e às preferências de cada cliente.

Construir relacionamento contínuo e de confiança

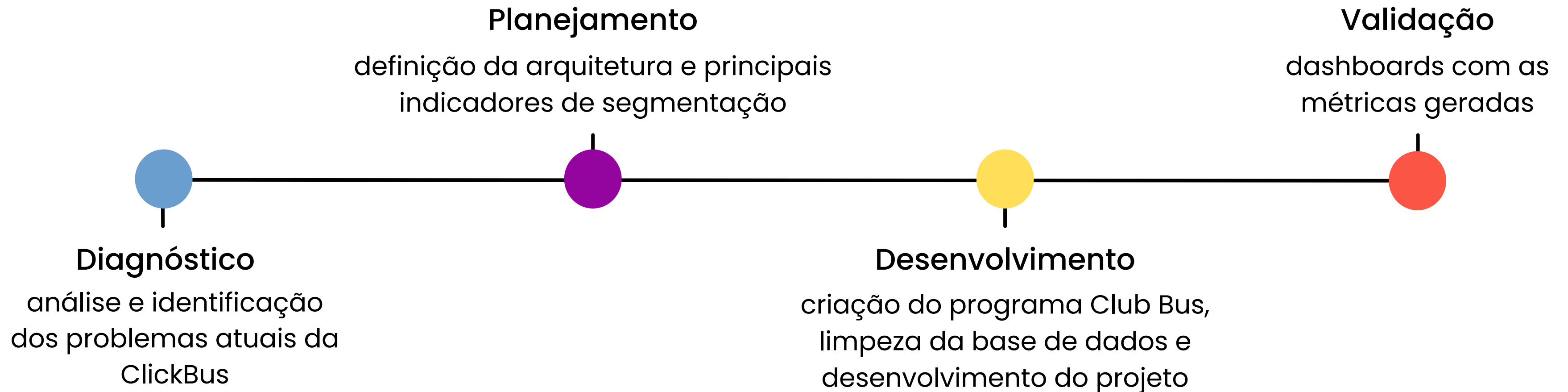
Ao entregar **experiências mais alinhadas ao perfil de cada usuário**, fortalecemos o vínculo com os clientes e incentivamos a fidelização dentro do Club Bus.

Proposta de Solução

Abordagem adotada

Optamos por uma solução baseada em dados, com **foco em segmentação** de clientes e **previsibilidade de comportamento**.

Etapas realizadas





Principais Funcionalidades

Segmentação dos Clientes

Agrupamento por perfis diversos baseado em métricas diferentes.

Personalização da Experiência

Site adaptado a partir do perfil e preferências de cada cliente.

Recomendações Personalizadas

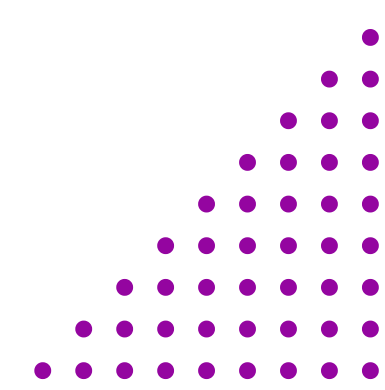
Sugestões de destinos, ofertas e campanhas adaptadas ao perfil de cada cliente.

Previsão de Rota

Identificação do próximo trecho provável com base em padrões históricos e contexto.

Previsão de Compra em 30 Dias

Estimativa da probabilidade de nova compra em janelas de tempo específicas.



Benefícios esperados:

Aumento da Fidelização com ações mais direcionadas

Segmentações precisas permitem campanhas personalizadas que fortalecem o vínculo e aumentam a recorrência.

Impulsionar o crescimento da plataforma com estratégia de dados

A ClickBus passa a tomar decisões mais inteligentes, antecipar demandas e ativar clientes com base em comportamento real.

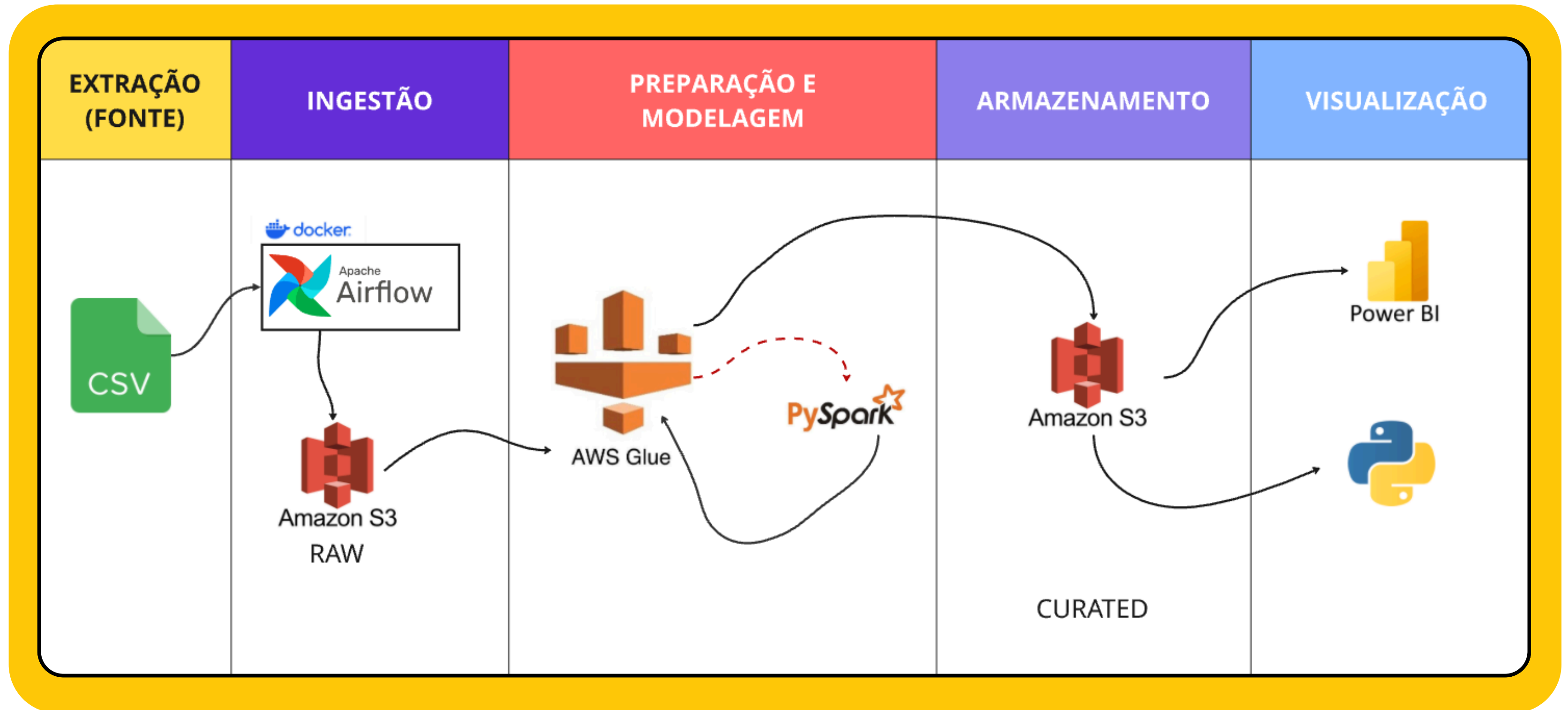
Melhoria na experiência do cliente com personalização

Ofertas e sugestões relevantes tornam a jornada mais fluida, prática e satisfatória para cada perfil.






Criar Vantagem Competitiva ao integrar inteligência com relacionamento




Ao unir previsões com engajamento contínuo, a ClickBus se diferencia no mercado com uma experiência única.

Arquitetura Utilizada



Tecnologias Utilizadas

TECNOLOGIA UTILIZADA	DESCRIÇÃO E USO NA SOLUÇÃO	VANTAGENS
	<ul style="list-style-type: none">• Fonte de dados principal fornecida pela ClickBus	<ul style="list-style-type: none">• Formato leve e universal• Fácil de manipular
	<ul style="list-style-type: none">• Plataforma de containerização - tudo o que a aplicação precisa para funcionar• Utilizado na solução para imagens do Airflow e MongoDB• Empacota código, dependências e configs juntos• Garante execução consistente em qualquer ambiente	<ul style="list-style-type: none">• Portabilidade• Isolamento de ferramentas• Rápido para subir e reiniciar serviços
	<ul style="list-style-type: none">• Orquestrador de workflows e tarefas ETL• Lê o arquivo .csv da pasta de entrada• Define fluxos como DAGs em Python	<ul style="list-style-type: none">• Controla o fluxo• Gratuito• Escalável e fácil de reproduzir entre ambientes
 Amazon S3 RAW E CURATED	<ul style="list-style-type: none">• Armazenamento de objetos escalável na nuvem• O Airflow salva no S3 os dados extraídos e pré-processados• Guarda dados brutos (<i>raw</i>) e tratados (<i>curated</i>) <p>MOTIVO SEPARAÇÃO (raw e curated):</p> <ul style="list-style-type: none">• Foi <u>solicitado pelo professor</u> de Big Data (Rogério) que utilizássemos um armazenamento Cloud e realizássemos as transformações dos dados, neste mesmo armazenamento (S3)• Dado isso, o primeiro arquivo ingerido estaria Raw, ou seja, do jeito que veio; Já o segundo, após transformações, estaria Curated, ou seja, organizado e processado.	<ul style="list-style-type: none">• Disponibilidade para dados brutos (<i>raw</i>) e tratados (<i>curated</i>)• Integração direta com Glue
 AWS Glue	<ul style="list-style-type: none">• Serviço ETL serverless da AWS• Permite combinar transformações SQL e PySpark• Unir consultas SQL com big data via PySpark <p>MOTIVO DO USO:</p> <ul style="list-style-type: none">• Conforme solicitação pelo professor Rogério, foi necessária a implantação de um serviço que realizasse transformações no armazenamento Cloud <p>MOTIVO DA TROCA (EC2 --> Glue):</p> <ul style="list-style-type: none">• Recomendação do professor Rogério;• EC2 é mais funcional para python e Athena para SQL, porém o Glue se conecta com ambos	<ul style="list-style-type: none">• Integração com serviços AWS• Flexível para unir consultas SQL rápidas com big data via PySpark

	<ul style="list-style-type: none">• Interface em Python para trabalhar com o Apache Spark• Processa grandes volumes de dados de forma distribuída.• Ideal para transformações complexas e de alto desempenho.• Muito usado em ETL, análise de big data e machine learning <p>PORQUE A MUDANÇA:</p> <ul style="list-style-type: none">• Conforme solicitado pela professora Patrícia, era necessária a utilização de Python para ler arquivos csv e realizar transformações e estatísticas sobre os dados. <p>PORQUE PYSPARK?</p> <ul style="list-style-type: none">• Dado que o professor de Big Data solicitou que a transformação e modelagem dos dados, fosse realizada durante o armazenamento em Cloud, foi necessária a utilização de Glue (para realizar ETL) e para conectar com o Python, houve a necessidade de PySpark.	<ul style="list-style-type: none">• Conexão com AWS S3 via AWS Glue• Escalabilidade automática para big data.• Suporte a múltiplas fontes e formatos.
	<ul style="list-style-type: none">• Linguagem de programação• Transforma dados em gráficos, facilitando insights.• Tem bibliotecas para visualização de dados• Faz conexão direta com MongoDB via pymongo.	<ul style="list-style-type: none">• Conexão com MongoDB• Domínio total sobre os dados• Suporte a bibliotecas• Documentação ampla
	<ul style="list-style-type: none">• Ferramenta de visualização e análise de dados.• Conecta com MongoDB• Exibe os resultados dos modelos• Visualização de dados prática• Permite storytelling	<ul style="list-style-type: none">• Conexão direta com MongoDB• Muito boa para gerar relatórios• Fácil uso e compreensão



Demonstração da Solução



1.Segmentação dos Clientes:

Segmentação Principal em Níveis de R\$ Gasto

Ao segmentar os clientes nestes grupos principais, o processo de identificação de perfis seria facilitado.

Realizamos segmentações por Feriados, datas comemorativas, festivais, finais de semana, entre outros.

Diamante			
999	6,512.95		
Total Clientes	Média Gasta (R\$)		
Platina			
11384	1,486.13		
Total Clientes	Média Gasta (R\$)		
Ouro			
30651	682.71		
Total Clientes	Média Gasta (R\$)		
Prata			
61867	350.91		
Total Clientes	Média Gasta (R\$)		
Bronze			
106257	163.82		
Total Clientes	Média Gasta (R\$)		
Novato			
77744	62.02		
Total Clientes	Média Gasta (R\$)		

Métricas utilizadas

Segmento	Total Clientes no Segmento	Média de Compras	Média Gasto R\$	Média de Tickets	Média Dias Ativos	Média Recência (dias)	Média Frequência Mensal	Média Ticket Médio R\$
Diamante	999	26.00	6512.95	38.18	22.72	897.36	33.17	396.99
Platina	11384	6.62	1486.13	9.76	6.39	1066.52	31.43	433.06
Ouro	30651	3.25	682.71	4.94	3.15	1278.84	31.20	341.70
Prata	61867	1.96	350.91	2.92	1.92	1482.77	30.86	243.42
Bronze	106257	1.34	163.82	1.77	1.32	1665.01	30.45	140.77
Novato	77744	1.08	62.02	1.22	1.07	1675.62	30.16	59.19

Novatos: menos de R\$100
Bronze: gastaram até 250

Prata: até R\$500
Ouro: até R\$1.000

Platina: até R\$3000
Diamante: mais de R\$4500

1.1: Extra - Dashboard sobre comportamento dos clientes



Segmentação dos Clientes por Nível

289 K

Total de Clientes

Distância percorrida por Segmentos (km)

Bronze

19.527.386.33

Novato

15.437.892,83

Prata

9.786.717.07

Oura

4.594.522,07

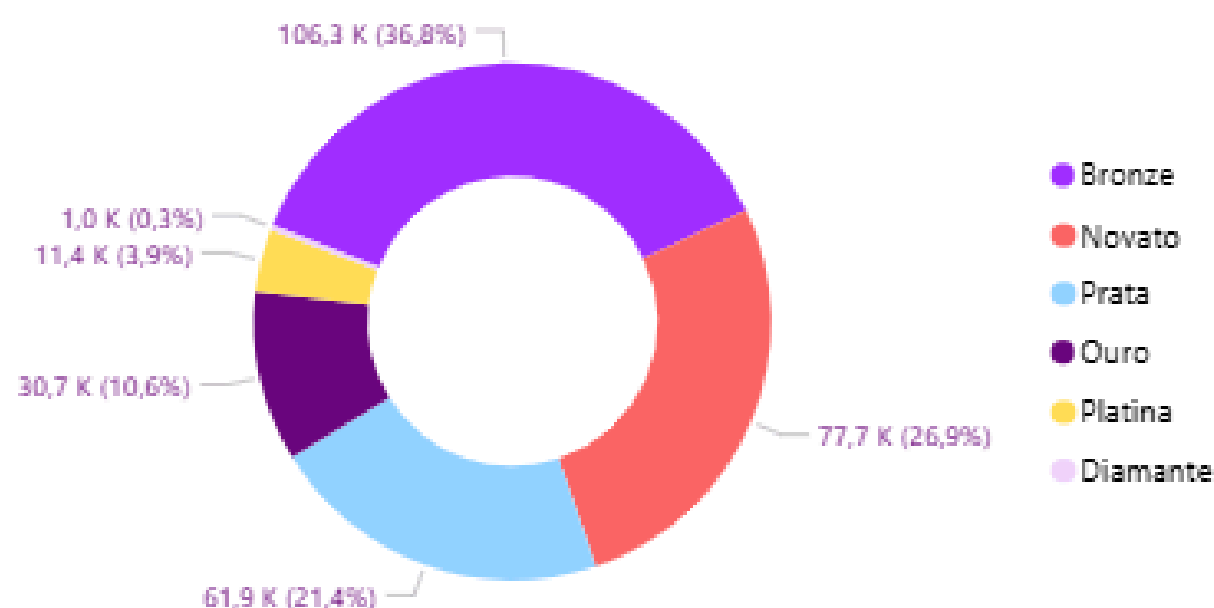
Platina

1.785.194,71

Diamante

143.628,89

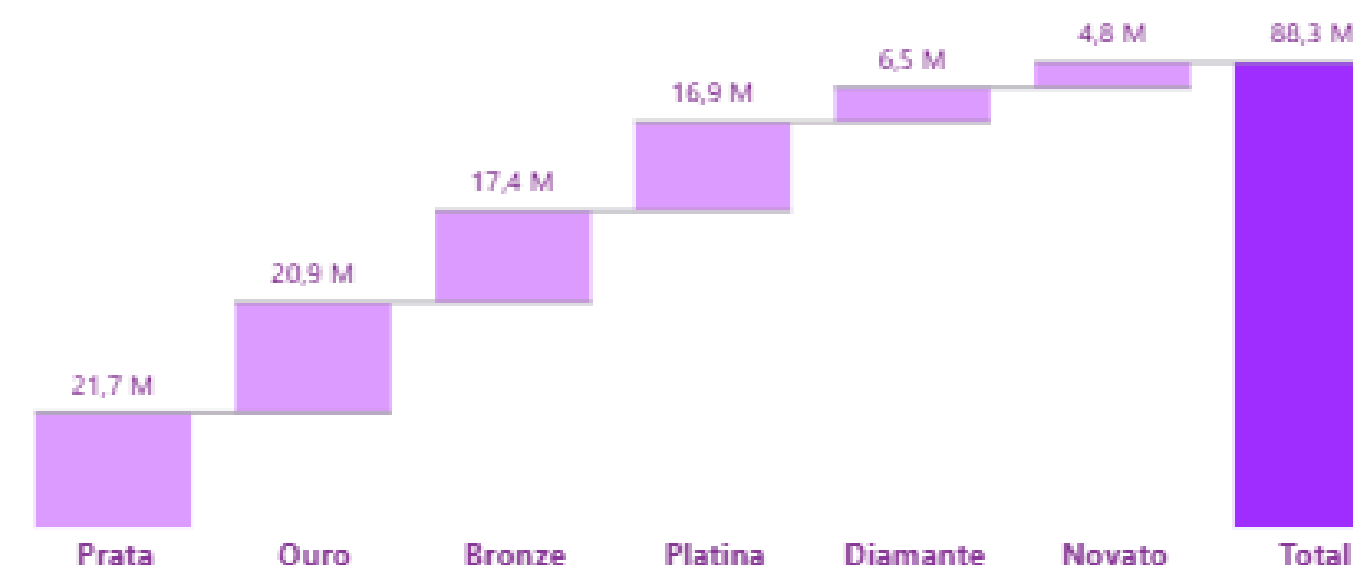
Cientes por Segmento



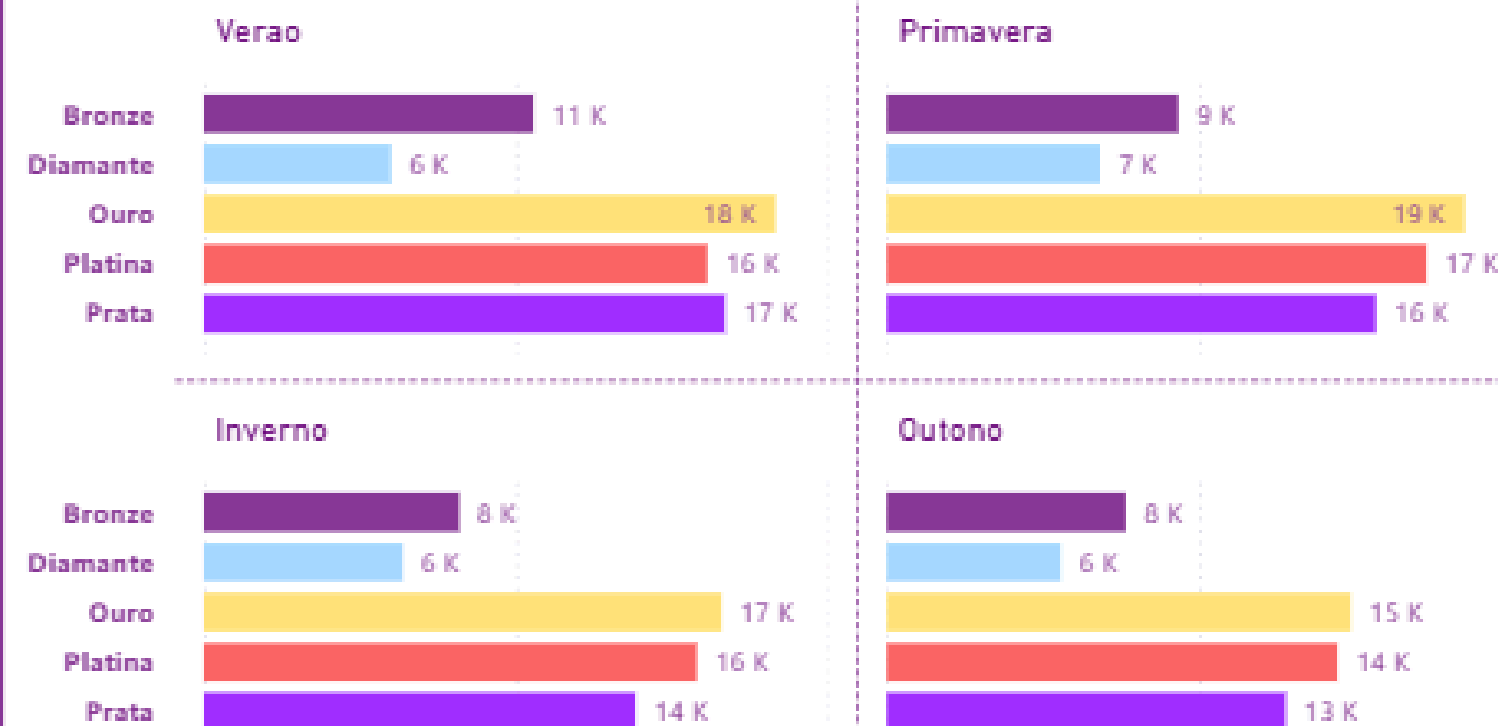
Compras por dia da Semana

Segmento	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
Ouro	18501	12277	8055	10550	8194	5550	5143
Platina	15392	11459	6457	9515	9026	6186	5600
Prata	17078	10941	8098	8675	6597	4011	3584
Bronze	10912	6448	5292	5812	3711	2077	1489
Diamante	8475	3828	903	2500	4070	2902	2091
Novato	2004	1141	1031	1030	566	300	166
Total	72362	46094	29836	38082	32164	21026	18073

Total Gasto por Segmento de Clientes



Compras por Estação do Ano





Segmentação dos Clientes por Nível

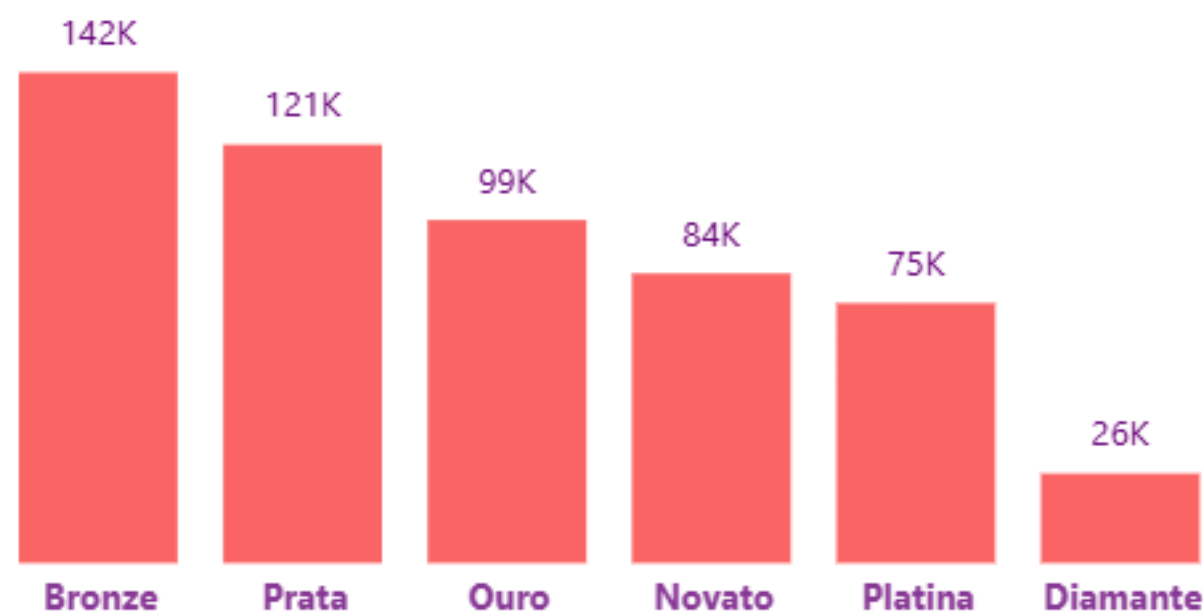
Métricas por Segmento

Nível	Frequência Média Mensal	Média de Dias Ativos	Média de Compras	Média de Tickets	Total Gasto
Prata	30.9	1.92	1.96	2.92	R\$ 21,709,776.95
Ouro	31.2	3.15	3.25	4.94	R\$ 20,925,609.00
Bronze	30.5	1.32	1.34	1.77	R\$ 17,407,403.60
Platina	31.4	6.39	6.62	9.76	R\$ 16,918,156.05
Diamante	33.2	22.72	26.00	38.18	R\$ 6,506,441.57
Novato	30.2	1.07	1.08	1.22	R\$ 4,821,479.58
Total	30.6	1.85	1.90	2.65	R\$ 88,288,866.75

Evento Preferido por Segmento

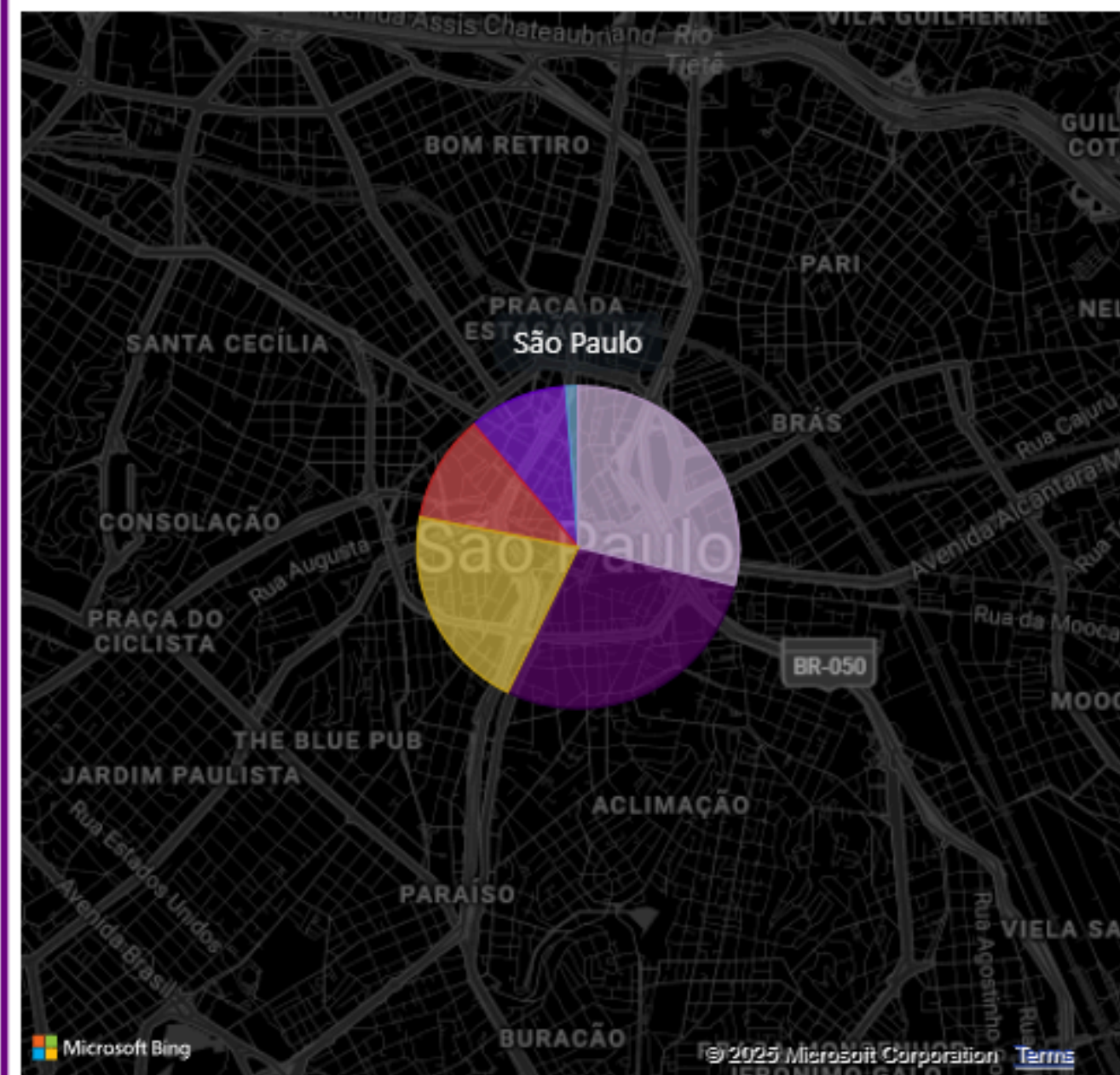
Bronze	Carnaval
Diamante	Natal
Novato	Finados
Ouro	Natal
Platina	Finados
Prata	Natal

Total Compras por Segmento



Principal Destino dos Segmentos

Segmento ● Bronze ● Diamante ● Novato ● Ouro ● Platina ● Prata

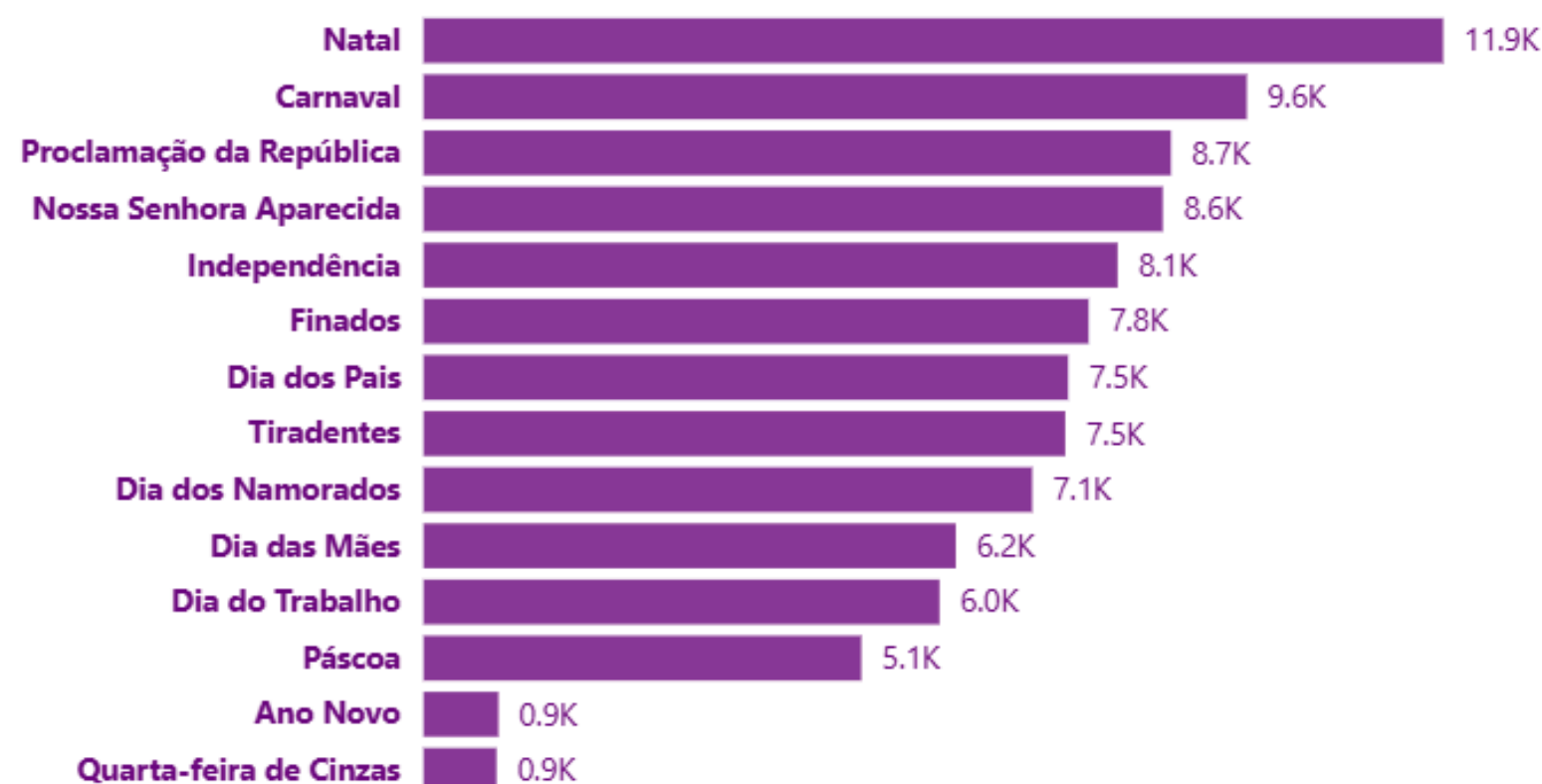


1.2: Dashboard sobre segmentação por métricas variadas



Segmentação por Métricas Variadas

Compras por Evento



Dia com Mais Compras

Friday

Dia com Menos Compras

Wednesday

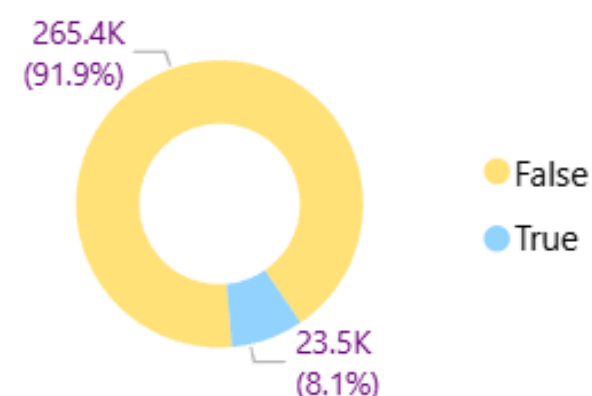
Cidades mais Visitadas

São Paulo	SP	37.23%
Rio de Janeiro	RJ	22.76%
Curitiba	PR	11.48%
Belo Horizonte	MG	8.47%
Brasília	DF	10.23%
Fortaleza	CE	9.83%

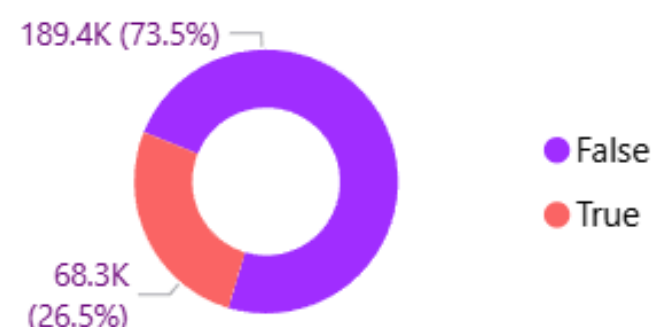
Cidades mais Visitadas



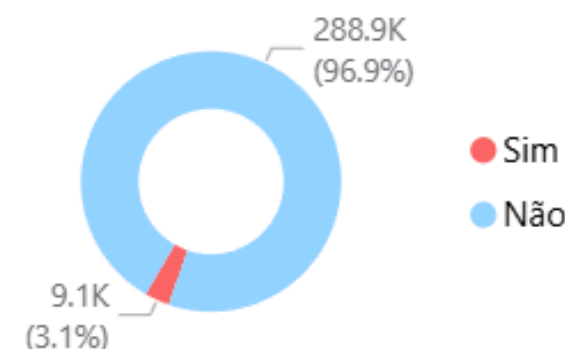
Cientes que Foram em Festivais



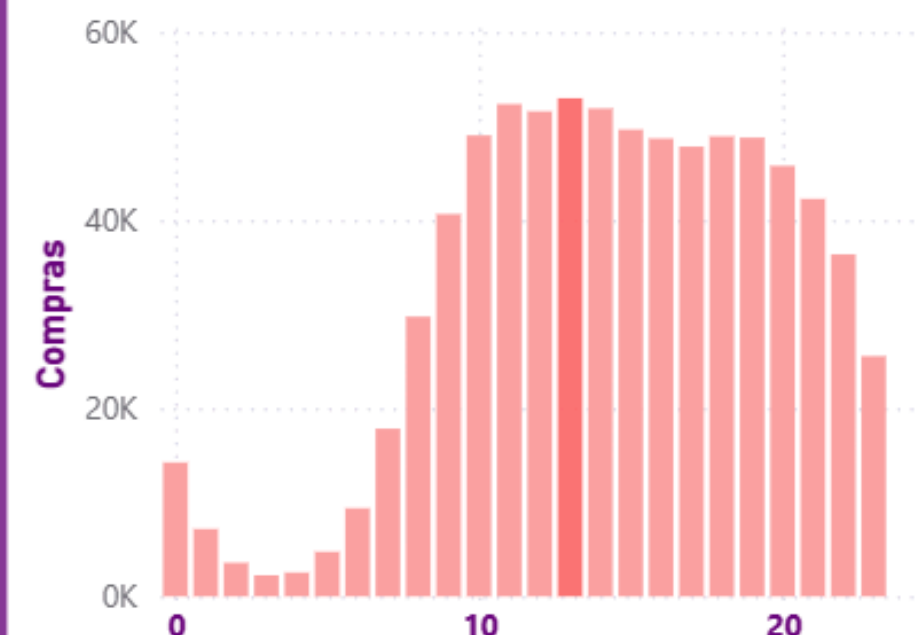
Cientes que Compram de Final de Semana



Cientes que Comprarão nos próximos 30 dias



Compras por Hora



2. Previsão de compra em 30 dias:

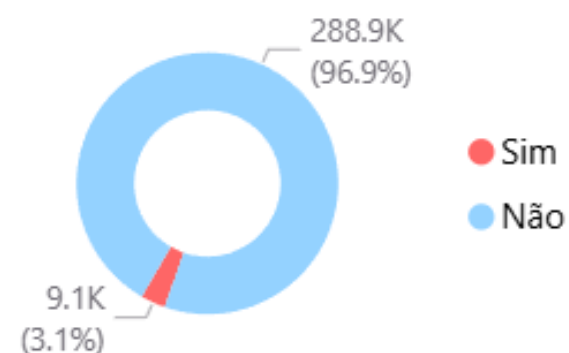
Método: Machine Learning

O modelo **Random Forest Classifier** apresentou excelente desempenho, com **acurácia de 99%** e **F1-score de média +90%**, confirmando a confiabilidade das previsões geradas.

	precision	recall	f1-score	support
0	0.993126	0.997712	0.995414	69943.000000
1	0.918367	0.788436	0.848456	2283.000000
accuracy	0.991097	0.991097	0.991097	0.991097
macro avg	0.955747	0.893074	0.921935	72226.000000
weighted avg	0.990763	0.991097	0.990769	72226.000000

AUC-ROC Score: 0.98

Clientes que Comprarão nos próximos 30 dias



Métricas utilizadas

- **Precision:** dos clientes previstos como compradores, quantos realmente compraram.
- **Recall:** dos clientes que compraram, quantos o modelo conseguiu identificar.
- **F1-score:** equilíbrio entre precisão e recall, garantindo que o modelo seja consistente.
- **Support:** quantidade de exemplos reais usados no teste.

Aplicação da Solução

- **Lembretes** sobre a Click Bus
- **Estímulos para a compra:** coupons se o cliente comprar em 48h por exemplo.
- **Sugestões e ações direcionadas** adaptadas ao perfil do cliente

3. Previsão da próxima rota:

Previsão da próxima rota

Utilizamos um modelo de **Logistic Regression** para prever a próxima rota de cada cliente com base no histórico de viagens.

```
Cliente selecionado aleatoriamente: 67868
      rota  probabilidade
Belo Horizonte -> São Paulo      0.9253
Belo Horizonte -> Rio de Janeiro  0.0621
Brasília -> Rio de Janeiro      0.0091
```

Resultado de Métricas Avaliativas:

Além da probabilidade

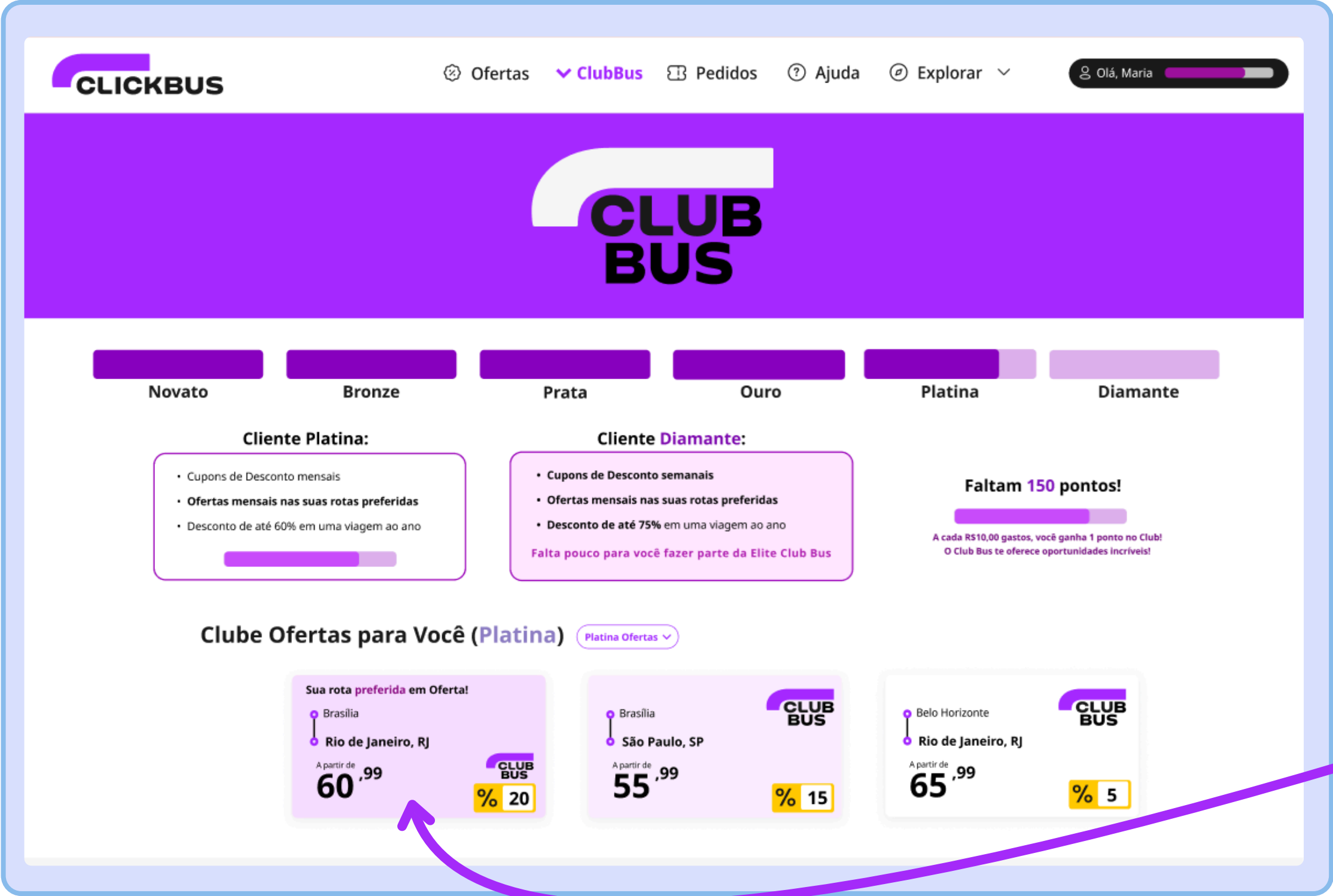
```
Acurácia: 0.98
F1 score: 0.98
```

Outros Testes (outro cliente):

```
Cliente selecionado aleatoriamente: 314615
      rota  probabilidade
Fortaleza -> Campo Grande      0.9915
Recife -> Curitiba      0.0073
Campo Grande -> Fortaleza      0.0006
```

Neste caso, é praticamente certo afirmar que o cliente irá realizar a rota de Fortaleza até Campo Grande.

3.1 Extra: aplicação da solução no site



Aplicação do Modelo

Com base nas previsões de rotas mais prováveis, o **cliente recebe** em sua área exclusiva do Club Bus **ofertas personalizadas** para suas próximas viagens.

Rotas mais prováveis:

```
Cliente selecionado aleatoriamente: 24065
                                     rota  probabilidade
Brasília -> Rio de Janeiro          0.8069
Brasília -> São Paulo               0.1221
Belo Horizonte -> Rio de Janeiro    0.0281
```

3.2 Junção da próxima rota + compra em 30 dias

Junção da próxima rota com a probabilidade do cliente comprar em 30 dias

```
Cliente 359369:  
Probabilidade de compra em 30 dias = 0.700 → SIM
```

```
Top 3 rotas prováveis:
```

	rota	probabilidade
	São Paulo -> Rio Verde	0.916
	São Paulo -> Rio de Janeiro	0.029
	São Paulo -> Nova Andradina	0.025

Próxima Compra + Rota Provável

Combinamos a **previsão de probabilidade de compra em até 30 dias** com as rotas mais prováveis para cada cliente. Neste teste o modelo indicou **70% de chance de compra** com +90% de chance da primeira rota.

4: Extra - Recomendação de Rotas Novas

Método: Matriz de Semelhança

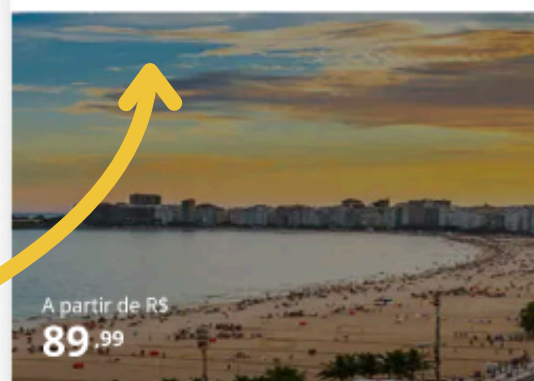
Aplicamos a técnica de **cosine similarity** para **sugerir rotas novas** a partir de **similaridade entre os clientes e seus perfis**.

Recomendações para o cliente 157068

```
['Bonito -> Rio de Janeiro',  
'Formosa -> São Paulo',  
'São Paulo -> Formosa',  
'Itacoatiara -> Formosa',  
'Formosa -> Itacoatiara',  
'Rio de Janeiro -> São Paulo']
```

Maria, que tal testar algo diferente?

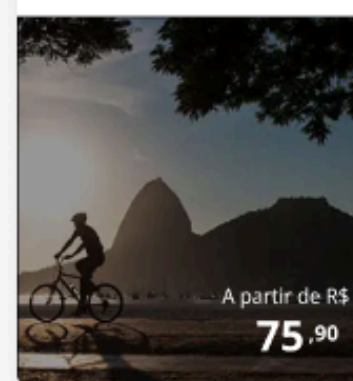
Bonito
Rio de Janeiro



Formosa
São Paulo



São Paulo
Formosa



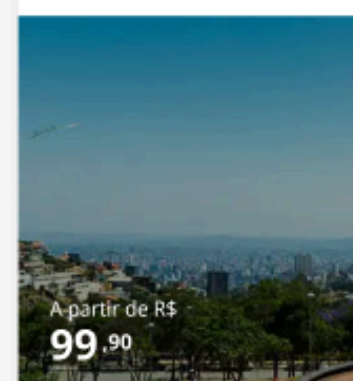
Itacoatiara
Formosa



Formosa
Itacoatiara



Rio de Janeiro
São Paulo



CLUB
BUS

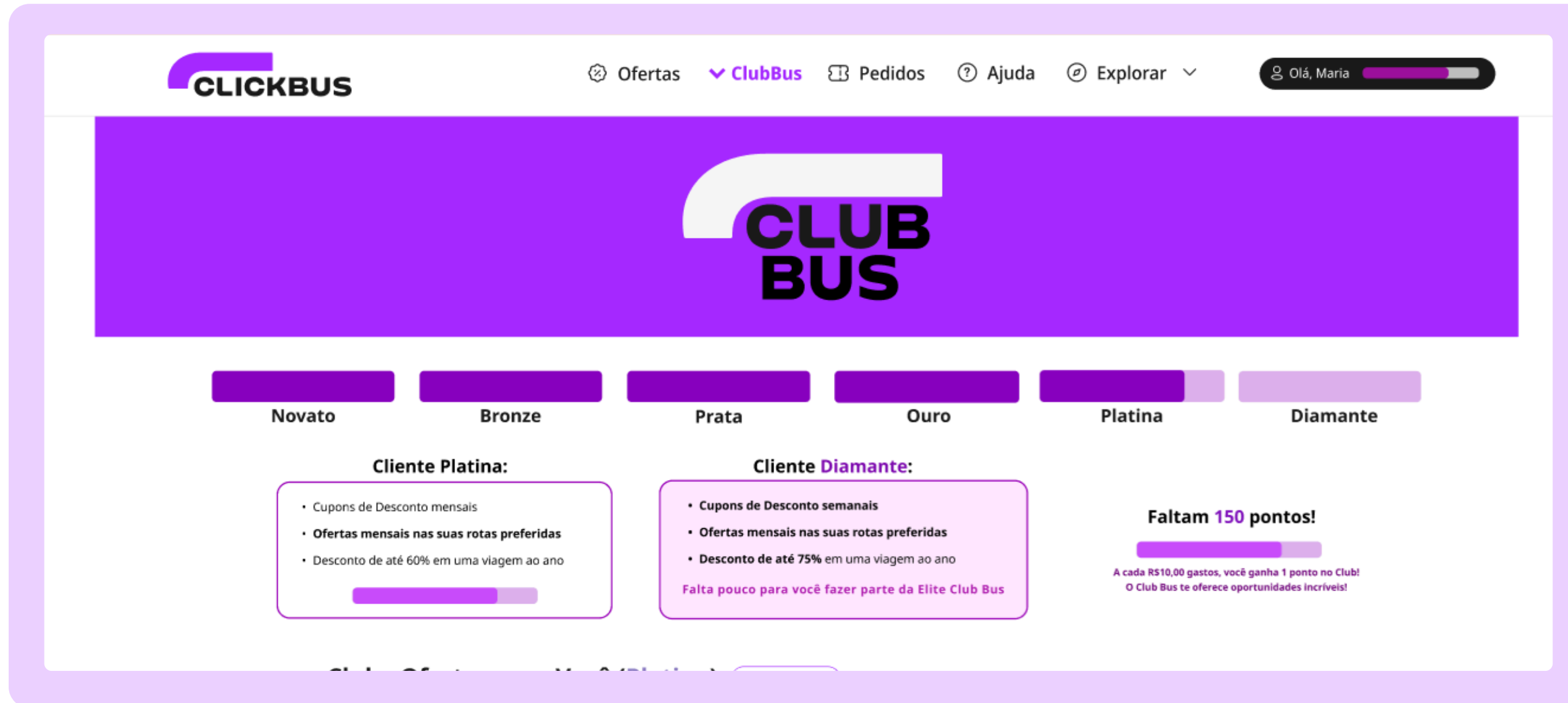
Quer saber como recomendamos
essas rotas para você?
Baseado no que você compra,
consequimos entender cada vez mais
seus gostos e preferências.

Queremos deixar sua experiência de
compra com a gente cada vez mais
especial.

Quer dar um feedback?

Queremos ouvir você

Site e E-mail Personalizado



As **previsões do modelo são aplicadas no site e nos e-mails**, oferecendo ofertas exclusivas e recomendações alinhadas ao perfil de cada cliente. Dessa forma, a ClickBus garante uma **comunicação personalizada** e aumenta as chances de conversão.





Dashboards Interativos



Dashboard Interativo

Ano
All

Cidade Destino
All

Cidade Origem
All

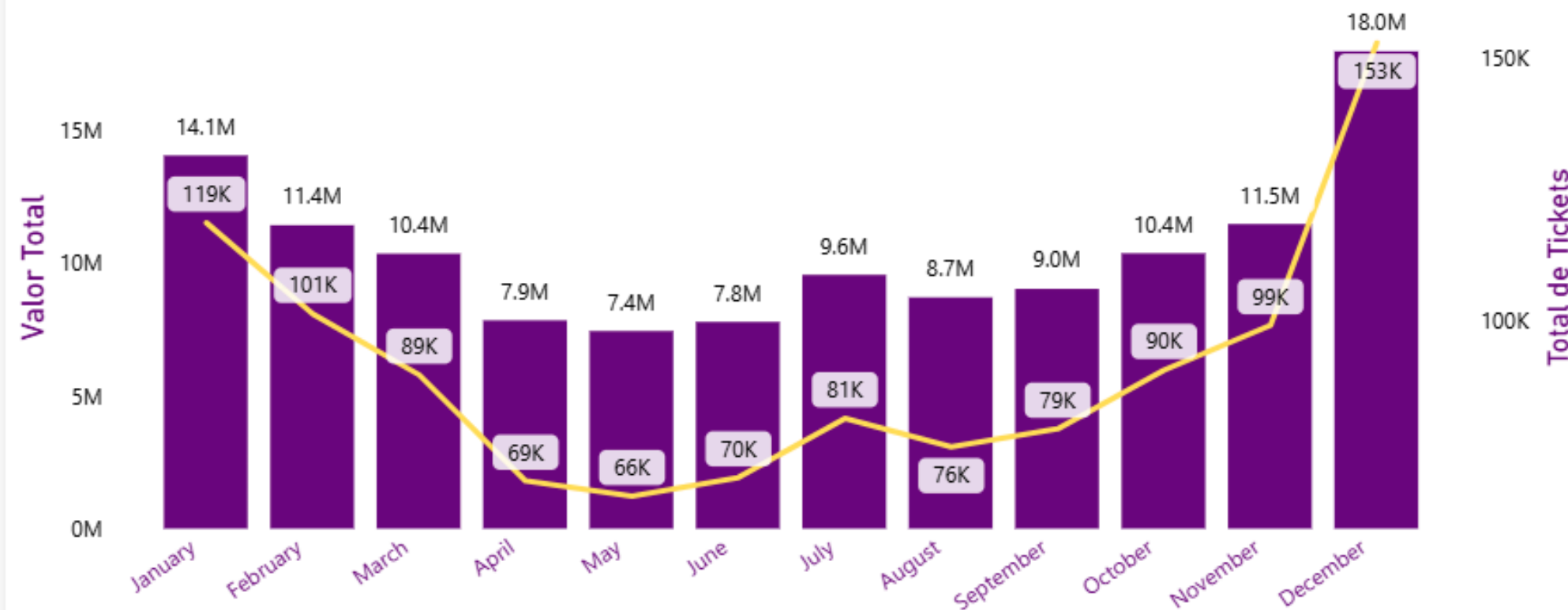
Estação do Ano
All

Dia da Semana
All

Remover Todos os Filtros

Histórico de Vendas

● Valor Total ● Total de Tickets



Vendas por Destino

Cidade Destino	Total Tickets	Valor Total
São Paulo	51378	R\$ 6,088,661.9
Rio de Janeiro	36129	R\$ 4,283,135.4
Curitiba	15171	R\$ 1,918,496.4
Belo Horizonte	12183	R\$ 1,746,270.2
Recife	9612	R\$ 1,356,558.7
Total	338694	R\$ 36,550,921.0

Vendas por Origem

Cidade Origem	Total Tickets	Valor Total
São Paulo	59372	7,201,723.76
Rio de Janeiro	39287	4,711,714.15
Curitiba	16419	2,038,784.61
Belo Horizonte	12409	1,791,642.81
Salvador	11254	1,634,882.75
Total	338694	36,550,921.00

Para visualização de outras métricas, criamos Dashboards Interativos, que permitem o uso de filtros para verificar informações relevantes, por data, estação, dia da semana e cidades.



Dashboard Interativo

Ano

All

Estação do Ano

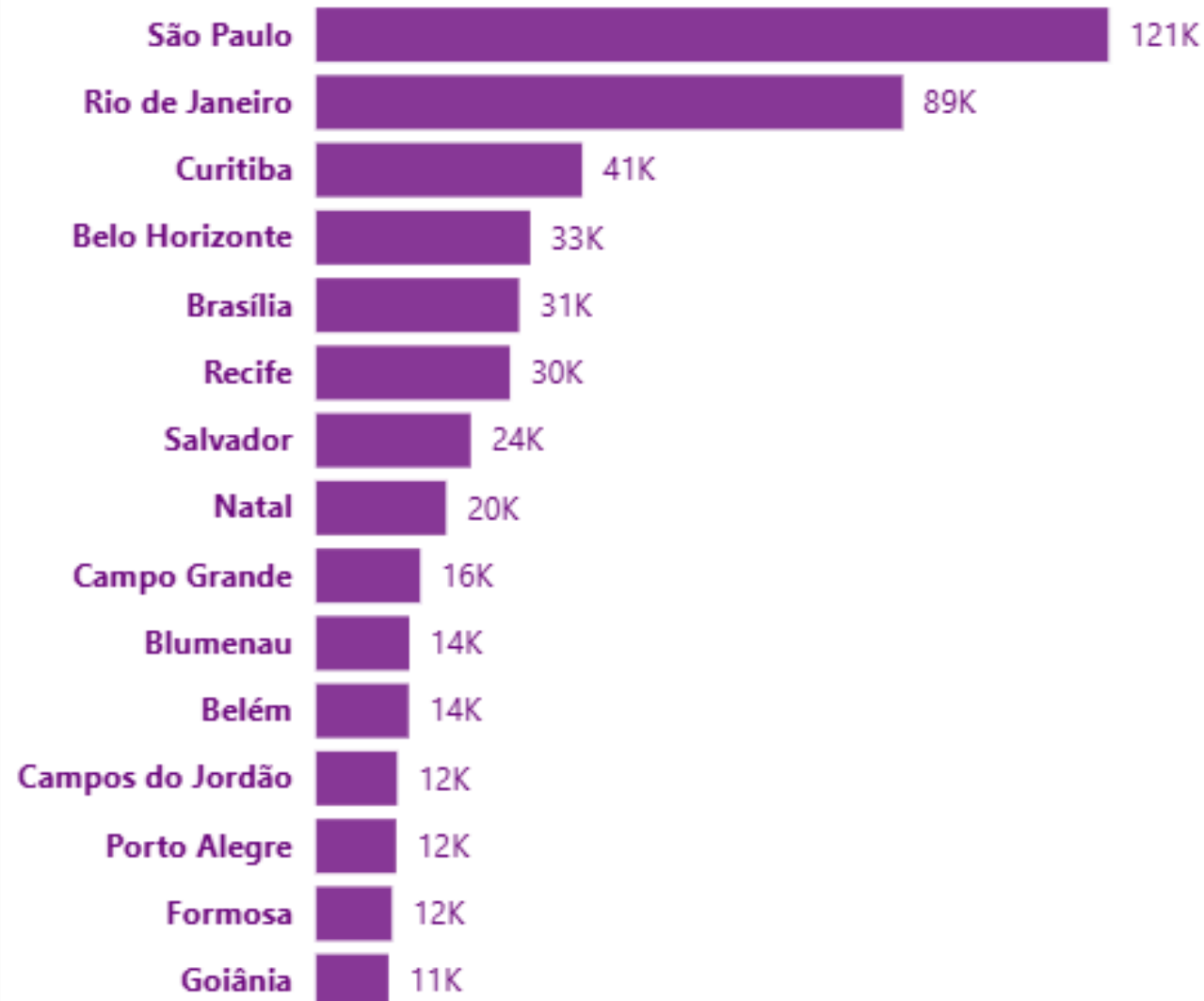
All

Dia da Semana

All

Remover Todos os Filtros

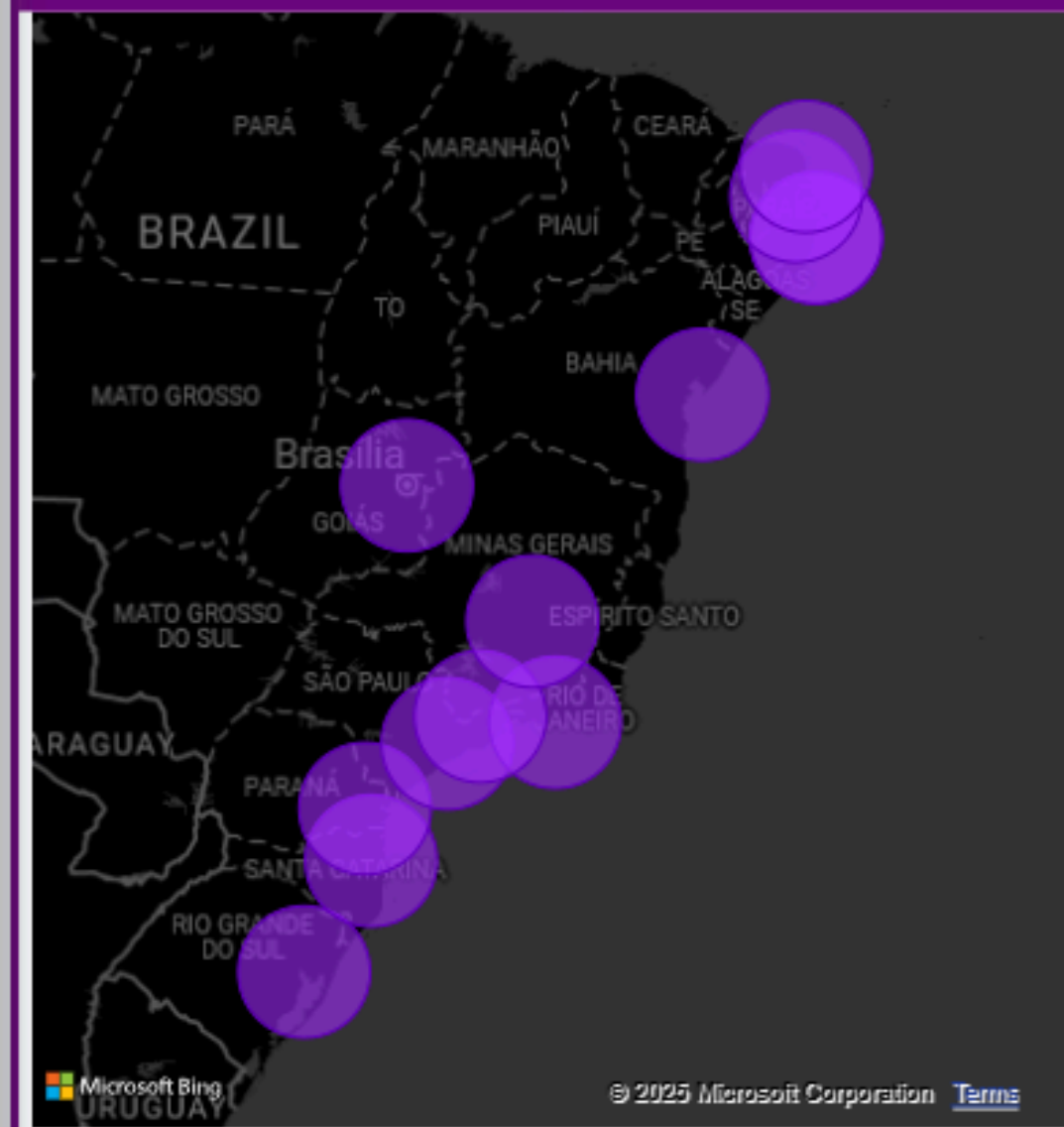
Tickets por Destino



Cidades que Mais Renderam (R\$)



Cidades mais Visitadas



Benefícios gerados

Para a ClickBus

Criar Vantagem Competitiva ao integrar inteligência com relacionamento

- Une previsões com engajamento contínuo e se diferencia no mercado com uma experiência única.

Campanhas e marketing com taxa maior de aproveitamento

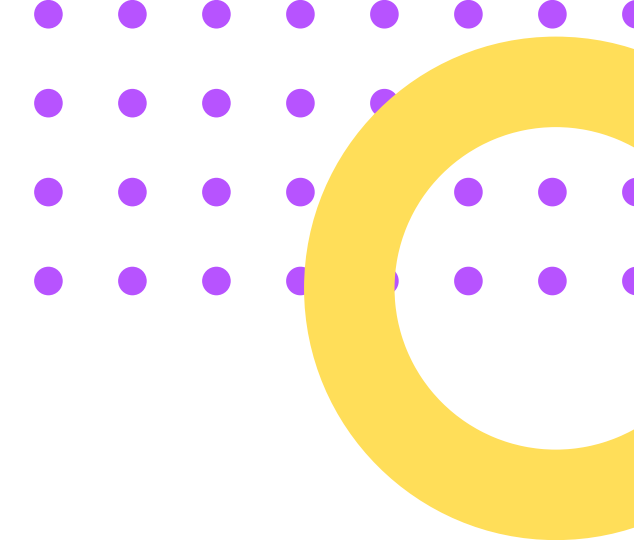
- Melhor aproveitamento dos investimentos em mídia

Aumento da Fidelização com ações mais direcionadas

- Campanhas personalizadas fortalecem o vínculo e aumentam a recorrência.

Personalização que converte interesse em compras

- Mais vendas com base em comportamento e preferências



Benefícios gerados

Para o Cliente

Tornar a experiência de compra mais rápida, eficiente e satisfatória

Menos tempo navegando, mais agilidade na compra e mais conexão com a empresa.

Personalização do site e e-mails

Conteúdo mais atrativo e direcionado, mantendo a satisfação do cliente

Recomendações personalizadas de destinos

Sugestões com base em perfis semelhantes, tornando a recomendação mais relevante

Previsões de Datas e Rotas

Ajuda o cliente com ofertas e planos já pensando nas possibilidades



Links Importantes

[Clique aqui e veja o PowerBI](#)

[Clique aqui e veja o Pitch](#)

[Clique aqui e veja o Protótipo](#)



Conclusão e Próximos Passos

Limitações Enfrentadas:

Os dados criptografados acabaram por tomar muito tempo da nossa solução, atrasando o processo de desenvolvimento.

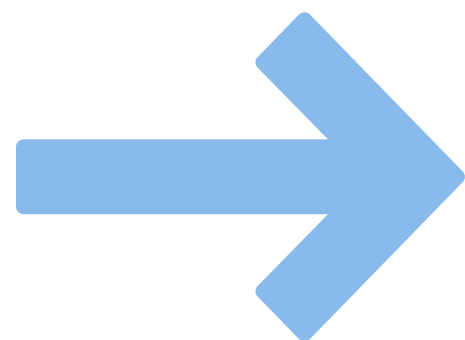
Conclusão

O desenvolvimento do projeto mostrou que tudo exige mais do que tecnologia... Entender o contexto do cliente e da empresa foi essencial para gerar uma solução efetiva.

Próximos Passos

Colocar em prática a solução nos sites e campanhas para testar a eficácia do projeto desenvolvido.

Ampliar análises preditivas com técnicas mais avançadas de IA e implementar testes A/B para otimizar campanhas personalizadas.



**AGRADECEMOS
PELA ATENÇÃO!**