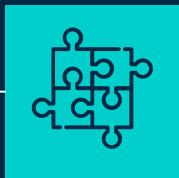


CASE: AliPaga

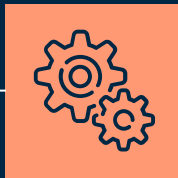
Análise, compreensão e
previsão

TABLE OF CONTENTS



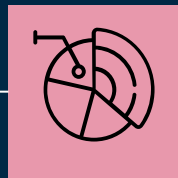
01

Análise e Compreensão dos
Dados



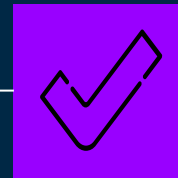
02

Previsão de Vendas



03

Relação dos Dados e
Recomendações



03

Conclusão

Análise e Compreensão dos Dados

01

Entendendo os Dados

- Possui 23 colunas com informações de compras realizadas por usuários do AliPaga por todo o mundo
- As colunas possuem formatos específicos e utilizadas para tipos diferentes de análises e gráficos
- Dataset com poucos “buracos”, só em postal code
- Clientes divididos em mais de 140 países
- Algumas colunas possuem relação de crescimento e decrescimento



Primeiras Observações

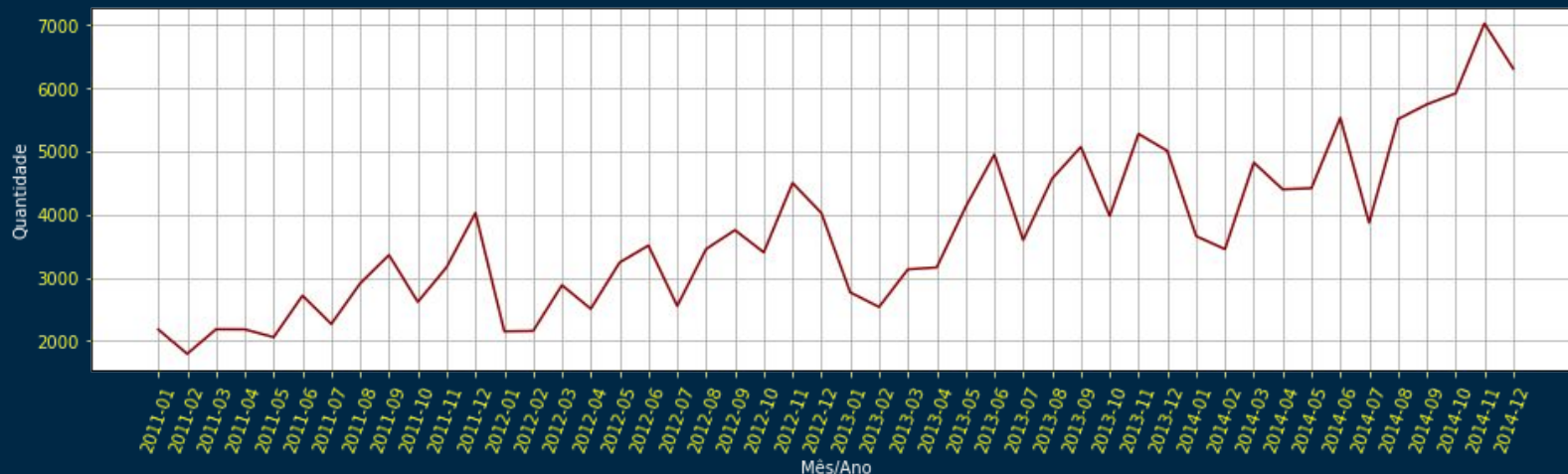
- AliPaga, no período de 4 anos (2011-2014) vendeu para clientes de 147 países distintos.
- Presenciamos um crescimento no valor total de vendas ao passar dos anos



Valor total de vendas por mês/ano

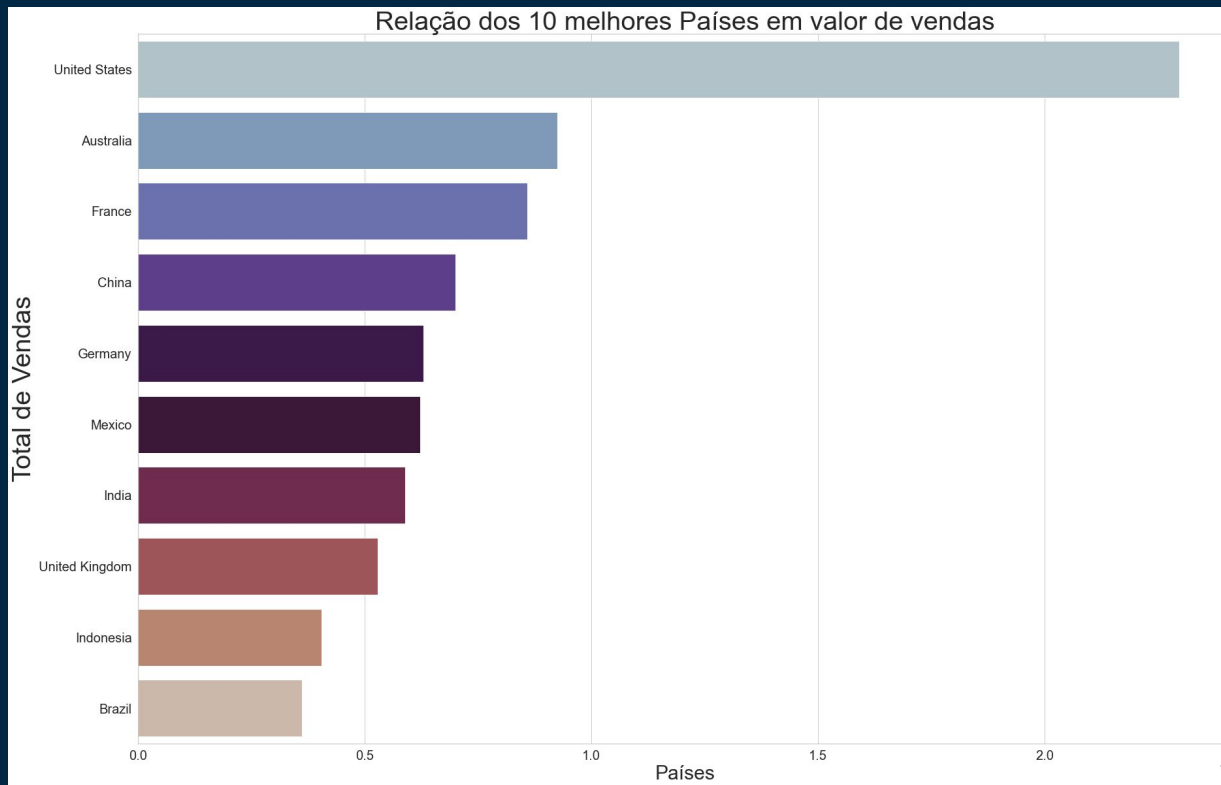
Primeiras Observações

- AliPaga, no período de 4 anos (2011-2014) vendeu para clientes de 147 países distintos.
- Presenciamos um crescimento no valor total de vendas ao passar dos anos



Quantidade de produtos vendidos por mês/ano

Maiores Consumidores



Maiores Consumidores

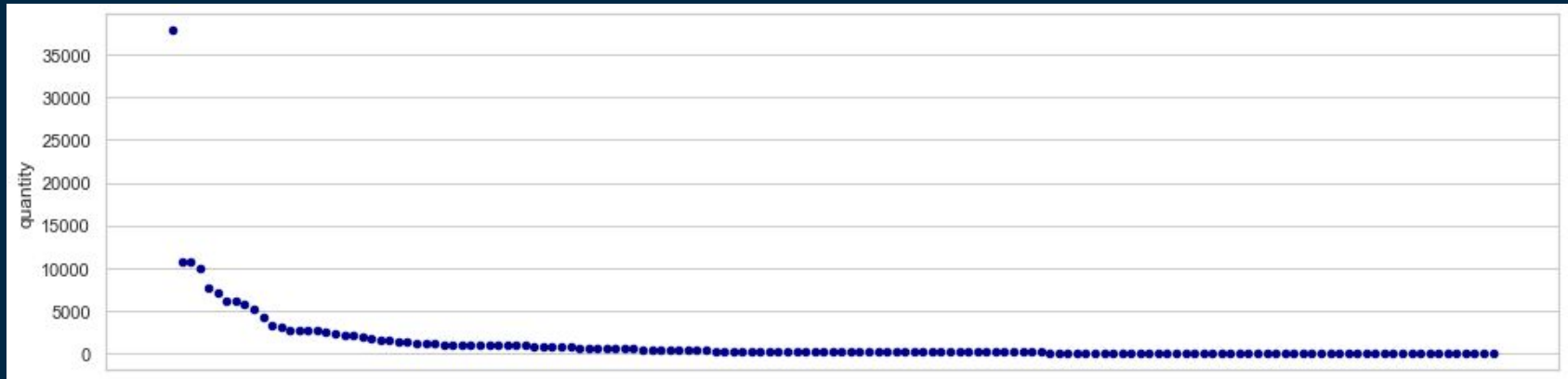
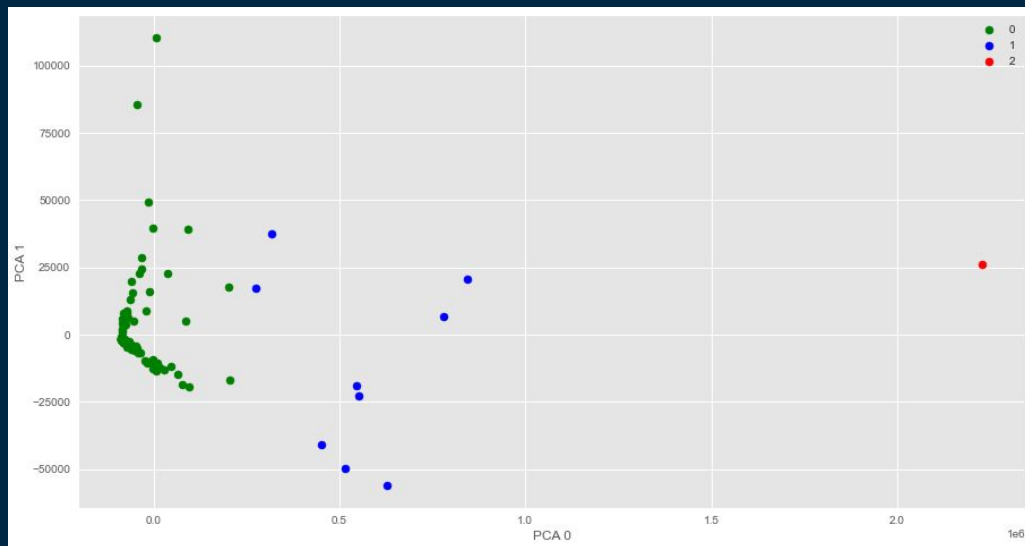


Gráfico dos países em relação a quantidade de vendas. Outlier = Estados Unidos

Maiores Consumidores - Clusterização

Foi utilizado o KMeans para uma simples clusterização dos países em função dos dados de Venda, quantidade de vendas e lucro.



country	cluster
United States	2
India	1
France	1
Mexico	1
China	1
Australia	1
Germany	1
United Kingdom	1
Indonesia	1
Brazil	1
Peru	0
Pakistan	0
Panama	0
Papua New Guinea	0

Plot PCA/Kmeans dos países em função de Vendas/Lucro.

Conclusões

- Os Estados Unidos são de fato a maior fonte de vendas do AliPaga no período em questão, com quase 4x mais o 2º lugar em quantidade de vendas (França)
- Isso poderia ser considerado como um outlier, porém não faria sentido remover ou regular esse item no contexto desse projeto
- O Algoritmo do Kmeans conseguiu fazer uma boa separação em 3 Clusters de países que são clientes do AliPaga, sendo no grupo 2 exclusivo para os Estados Unidos.

Previsão de Dados

02

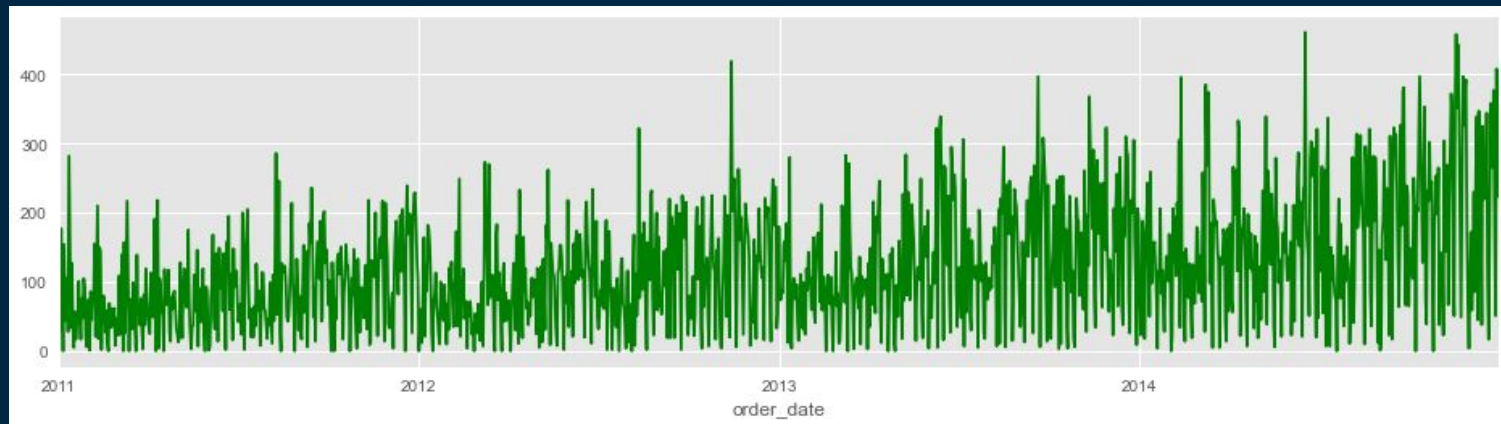
Forecast de Vendas

- Nessa seção o foco foi em realizar uma predição de valores de venda do “próximo” mês
- O período em questão a ser previsto é o período que corresponde ao primeiro mês do ano de 2015, que seria o próximo mês após o último dia disponível no conjunto de dados (31/12/2014)
- Nesses testes foram utilizados apenas a quantidade de vendas e não seu valor total



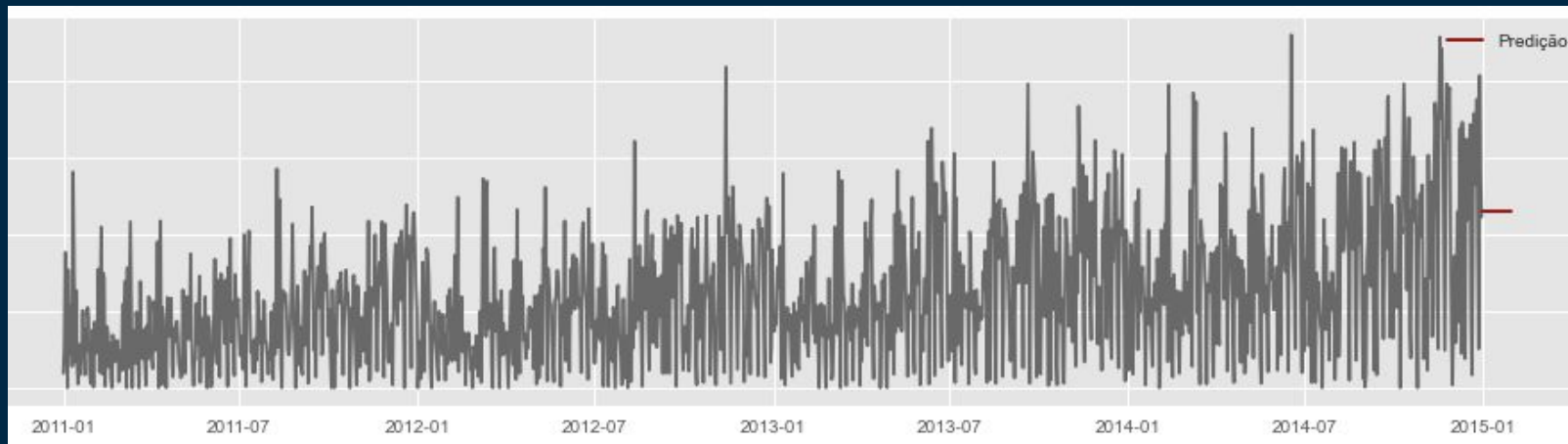
Série Temporal

- A série temporal de quantidade de vendas gerada pelo dataset possui algumas características:
 - É estacionária
 - Não possui tendências ou sazonalidade
- Os dias faltantes foram substituídos por valores nulos (0)



Série Temporal

- Foi utilizado todo o período de 4 anos para treino
- A partir do conjunto de teste, o objetivo é prever os próximos 31 dias pois o mês em questão é Janeiro/2015



Detalhes e Resultados

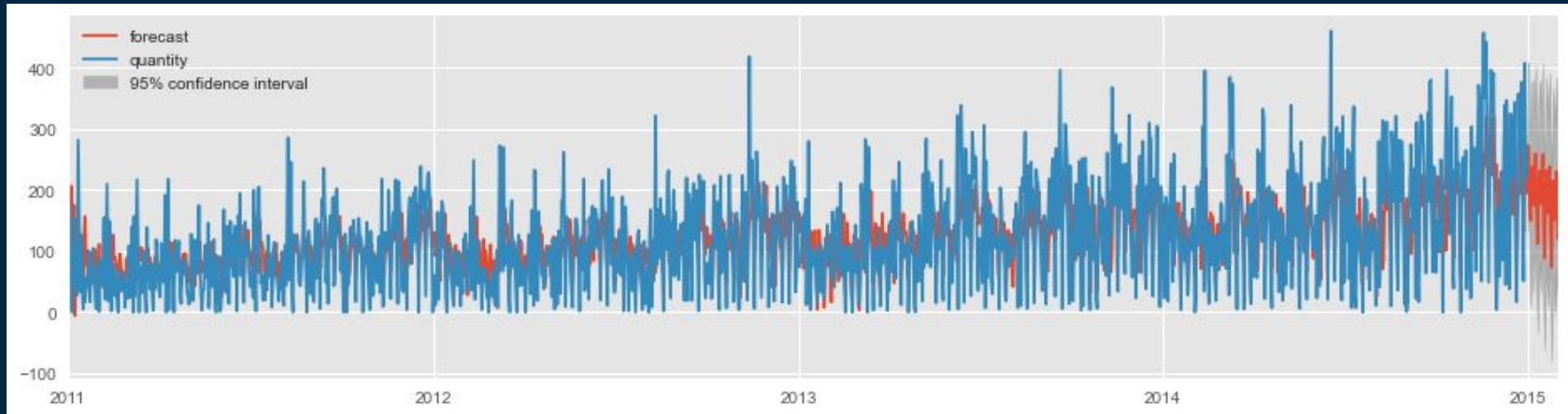
- O modelo utilizado foi ARIMA (Autoregressive moving average model), um modelo que utiliza autocorrelação e médias móveis para prever o futuro.

2015-01-01	273.0	2015-01-11	113.0	2015-01-21	163.0
2015-01-02	271.0	2015-01-12	202.0	2015-01-22	202.0
2015-01-03	178.0	2015-01-13	231.0	2015-01-23	238.0
2015-01-04	150.0	2015-01-14	178.0	2015-01-24	131.0
2015-01-05	229.0	2015-01-15	216.0	2015-01-25	73.0
2015-01-06	239.0	2015-01-16	259.0	2015-01-26	186.0
2015-01-07	176.0	2015-01-17	145.0	2015-01-27	216.0
2015-01-08	223.0	2015-01-18	89.0	2015-01-28	152.0
2015-01-09	260.0	2015-01-19	192.0	2015-01-29	186.0
2015-01-10	168.0	2015-01-20	232.0	2015-01-30	231.0
				2015-01-31	119.0

Valores previstos para 01/2015

Detalhes e Resultados

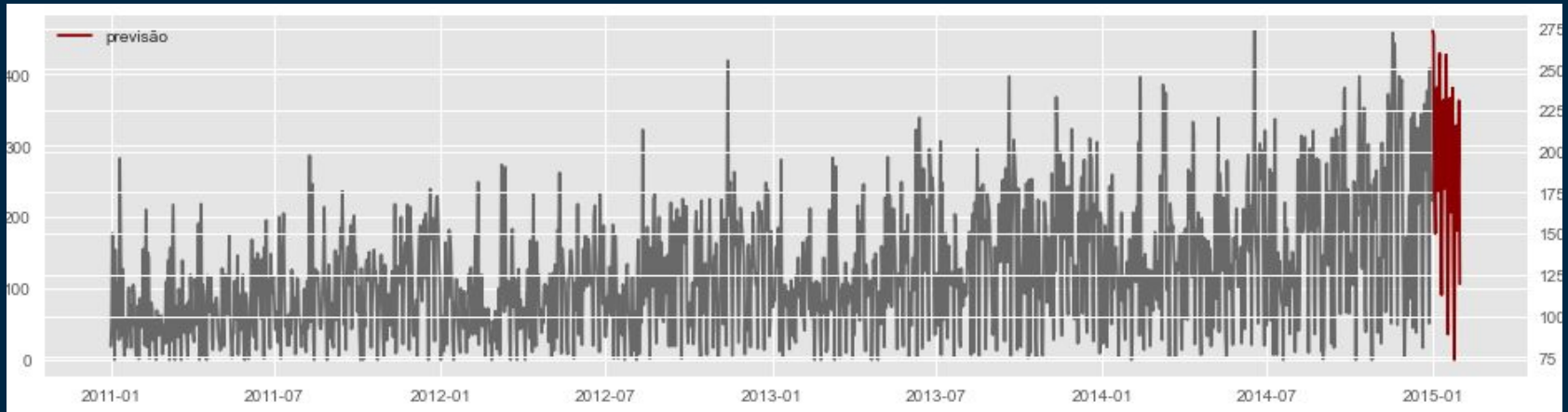
- A predição foi feita em todo o período de treino + o período de teste (01/2015)



Forecast de quantidades vendidas em todo o período de 01/2011 - 01/2015

Detalhes e Resultados

- A predição foi feita em todo o período de treino + o período de teste (01/2015)



Dados de quantidades vendidas em todo o período de 01/2011 - 12/2014 + Previsão de 01/2015

Conclusões

- Apesar de aparentes bons resultados, esse modelo possui um valor AIC (Informação de Akaike) alto o que pode significar perda de informação, então reduzir esse valor pode melhorar o resultado do forecast

Relação dos Dados, recomendações e conclusão

03

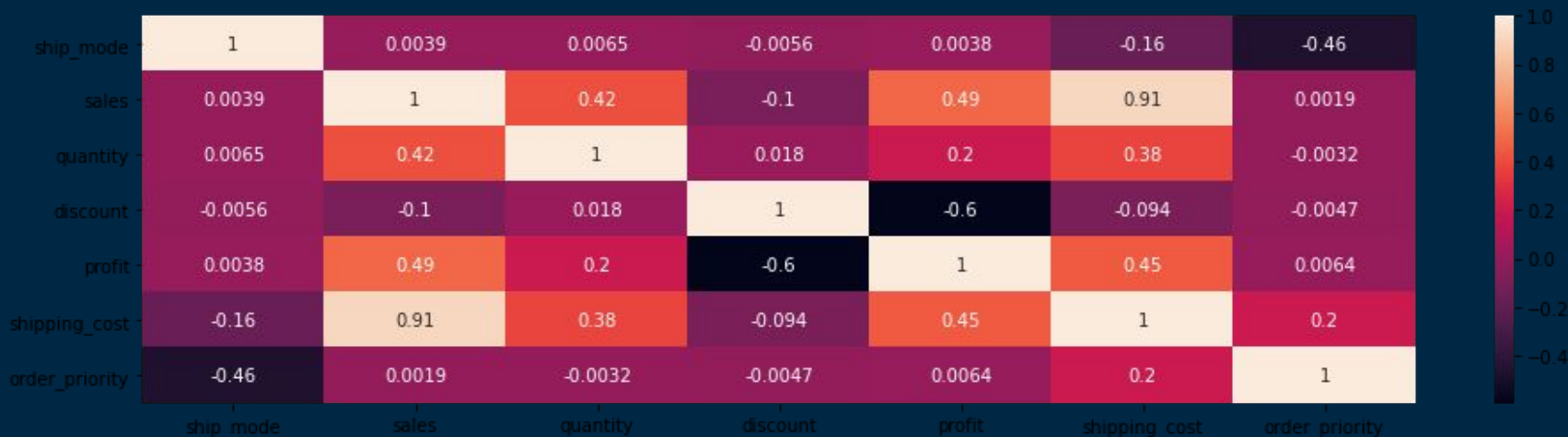
Shipping Cost e Discount

- Nessa seção será abordado o uso e a importância dessas duas colunas nos resultados (vendas, quantidade, etc.)
- Iremos ver a relação dessas variáveis em relação às outras presentes no nosso conjunto de dados

	\$25 \$5 shipping
	\$30 Free shipping

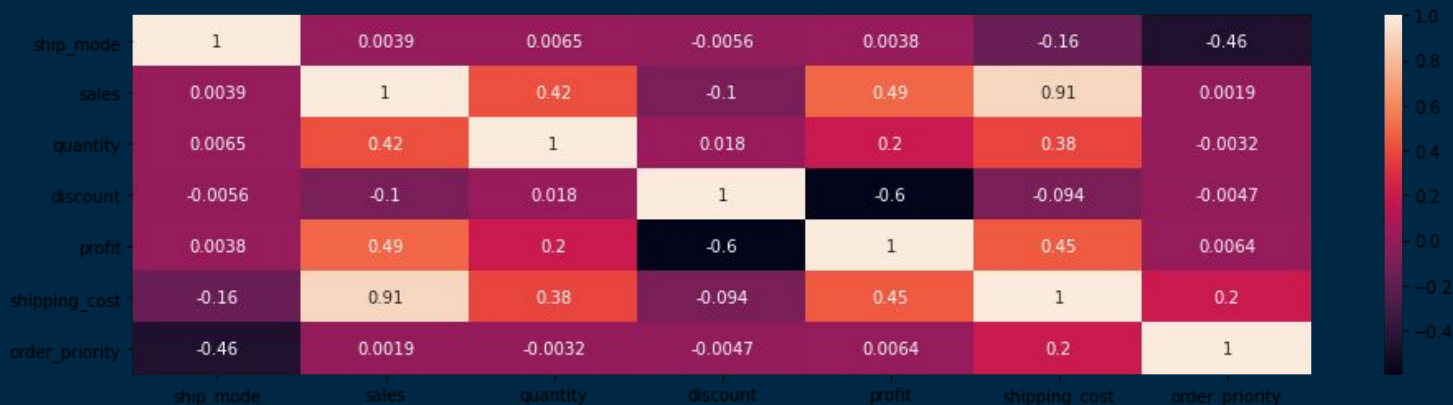
Shipping Cost e Discount

- Foi abordado principalmente o uso da correlação de spearman que avalia a relação monotônica entre duas variáveis
- A correlação entre as variáveis podem ser exibidas em um heatmap



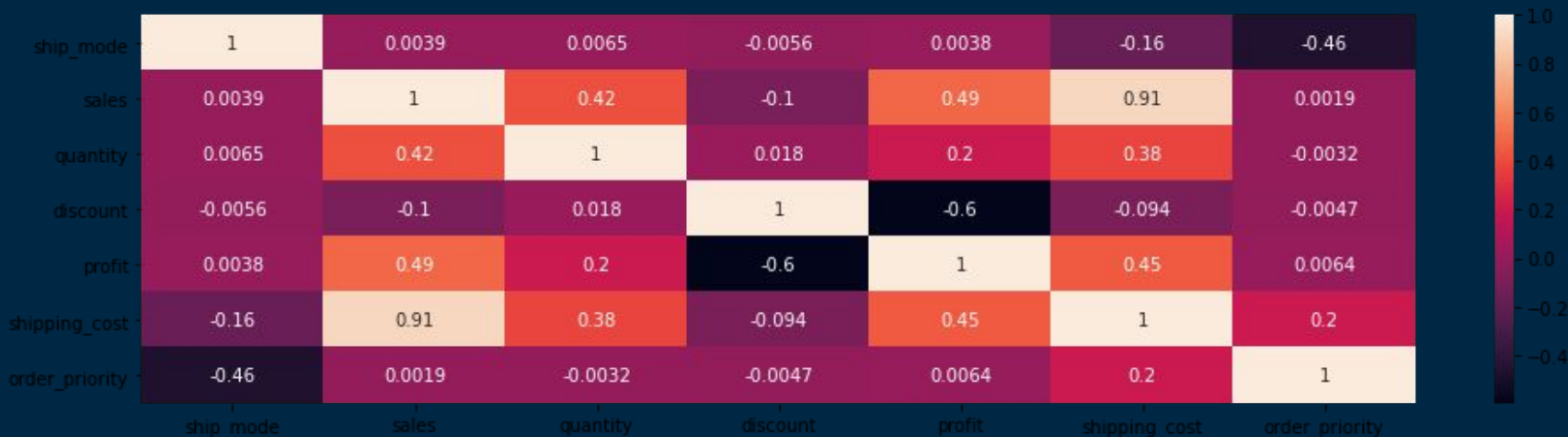
Sobre o Shipping Cost (Frete)

- Com a tabela acima podemos notar algumas coisas sobre o shipping_cost:
- Há uma relação fortíssima entre sales e shipping_cost, isso se deve ao fato do cliente claramente ter pago o valor de entrega de um produto e por isso impacta diretamente no valor da venda
- Temos uma relação moderada entre profit e shipping_cost, provavelmente também está relacionado ao fato do lucro ser maior conforme o preço de entrega cresce
- Há uma leve relação de crescimento entre shipping_cost e quantidade de itens vendidos, talvez seja um fator que não tenha muito impacto na quantidade de vendas.



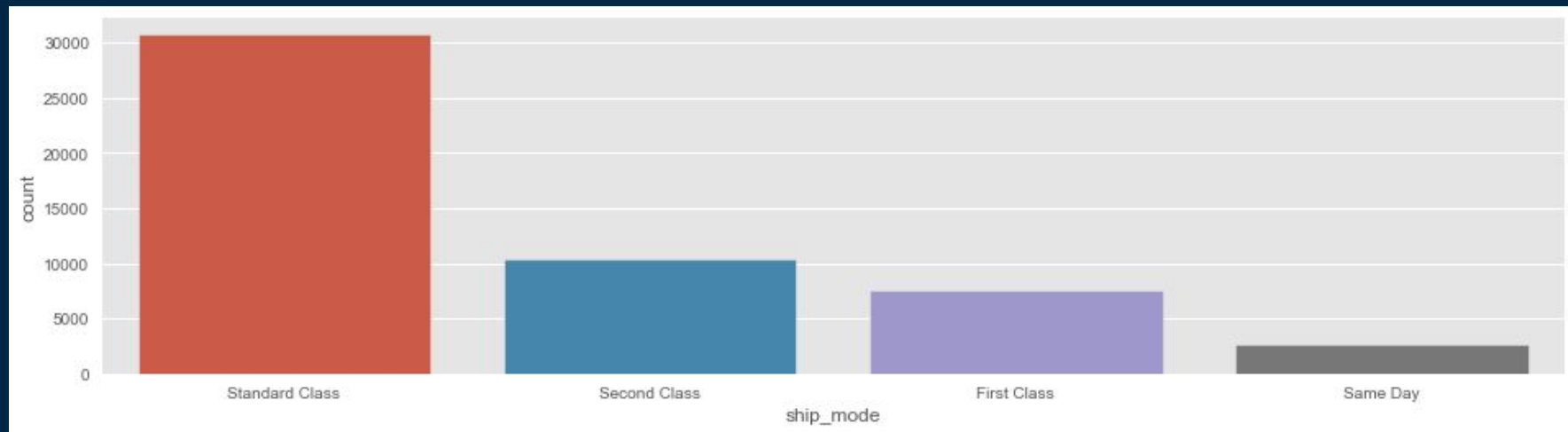
Sobre o Discount (Desconto)

- Podemos conferir que a relação discount/profit é fortemente negativa, o que quer dizer que o lucro de fato é menor conforme o desconto atribuído, o que faz sentido.
- No geral o fator desconto parece não impactar muito nos outros dados



Ship Mode (Modo de Envio)

- Foi realizada uma análise sobre a preferência dos modos de envio



Plot dos Modos de envio por quantidade

Ship Mode (Modo de Envio)

- Com os dados disponíveis podemos fazer a média dos valores das opções de envio

- Standard Class: 19.97
- Second Class: 30.46
- First Class: 41.05
- Same Day: 42.93

Recomendações

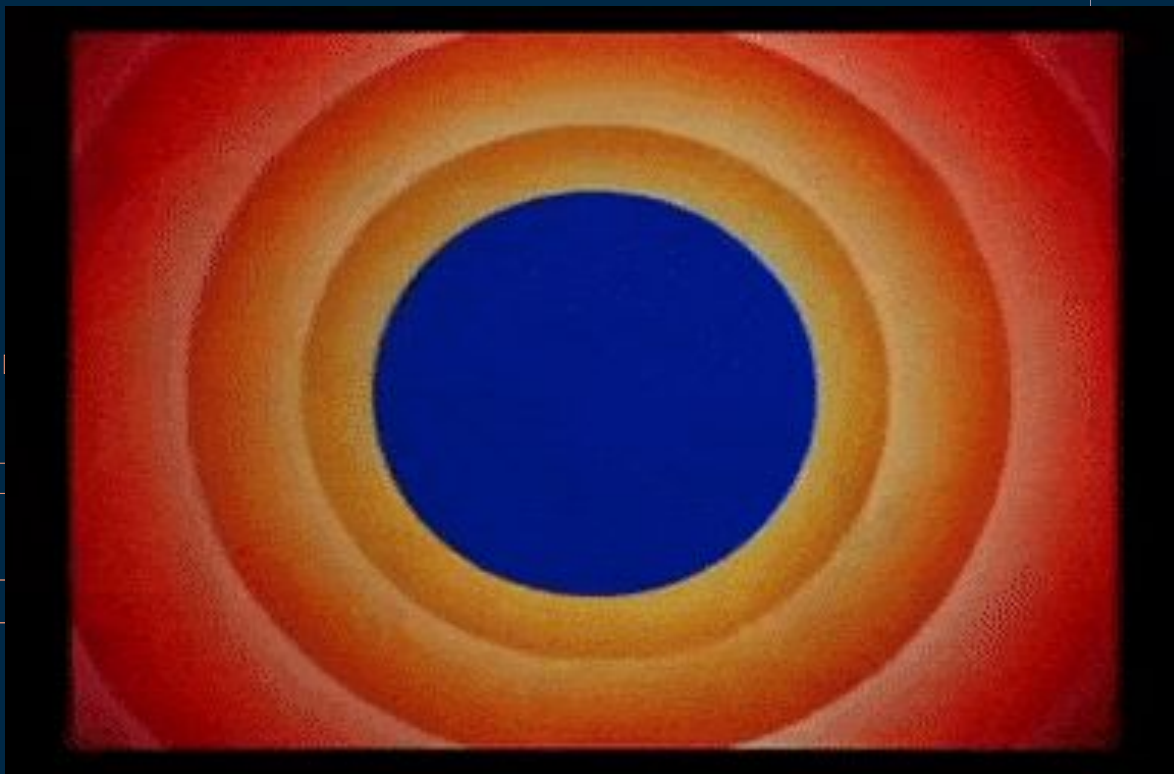
- Com o que foi mostrado é possível fazer algumas recomendações, entre elas:
 - Como apresentada, a relação entre shipping cost e as outras variáveis, como vendas, quantidade e lucro, esse pode ser o melhor ferramenta de lucro do AliPaga, porém pode haver consequências naturais como queda no número de usuários.
 - Há um relação inversa entre a prioridade do pacote o modo de envio, ou seja, quanto maior um é, menor o outro é, então caso um pacote seja muito importante ele provavelmente será enviado no mesmo dia, então aumentar o preço do "Same Day" pode ser benéfico.
 - O desconto em si não se relaciona muito com as outras variáveis, com exceção do profit que é inversamente proporcional ao seu crescimento, logo o lucro será maior quanto maior o desconto, o que é bem usual
 - Sobre o ship mode uma coisa a ser feita poderia ser a remoção da primeira classe e aumento do "Same Day", pois assim o usuário teria menos opção e pagaria mais nesse caso, pois os valores desses dois modos são muito próximos

Conclusão

04

Conclusão e Possíveis Melhorias

- Uma análise periódica de quantidade de compradores mensais
- Uma relação de valores de frete/usuarios/compras
- Um estudo mais profundo de como o desconto afeta a transação do cliente
- Melhorias no forecasting



CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#)