



Teste Cientista de Dados com foco em Growth - OLIST

Robson Müller - 29/07/20

Metodologia

Descritivo

- Para este desafio e qualquer outro em negócios, busco me basear em 3 frentes; utilizar a metod. data science unido à minha experiência em business e o método científico em si.
- Meu raciocínio fluirá desta forma:
 1. Entendimento do Negócio
 2. Problemas a serem resolvidos
 3. Requisitos, Levantamento e entendimento dos dados
 4. Transformação dos dados
 5. Modelagens e Hipóteses
 6. Validação e Estratégia
 7. Deploy de algum modelo
 8. Avaliação/Feedback

Os dois notebooks contendo a estruturação das análises estão no mesmo diretório desta apresentação.

Entendimento do Negócio

- Estamos interpretando o negócio como sendo a área de atuação no time de Marketing
- Acreditamos que a necessidade seja gerar valor com dados estruturados e não estruturados de n fontes, para suportar experimentos e validar hipóteses

Problemas a serem resolvidos

- Ingerir os dados do banco afim de gerar valor e insights que possam suportar experimentos na área de marketing focado em incremento de conversão e/ou receita.
 - Exemplo:
 - Aumentar a satisfação dos clientes aumentaria a conversão ?
 - Qual categoria de produto é a mais problemática ?
 - Poderíamos prever clientes insatisfeitos e realizar testes com eles ?
 - Que ganho teríamos acelerando o processo de descoberta de clientes insatisfeitos ?

Análise dos Reviews

- Seguimos a primeira dica do teste e comecei pela análise dos comentários, depois de entender o data schema, fiz alguns tratamentos, entre eles:
 - Comentários e títulos preenchidos em campo de data;
 - Gravar comentarios e títulos de review nos campos corretos;
 - Etc;
 - Dados filtrados, iniciamos a exploração;

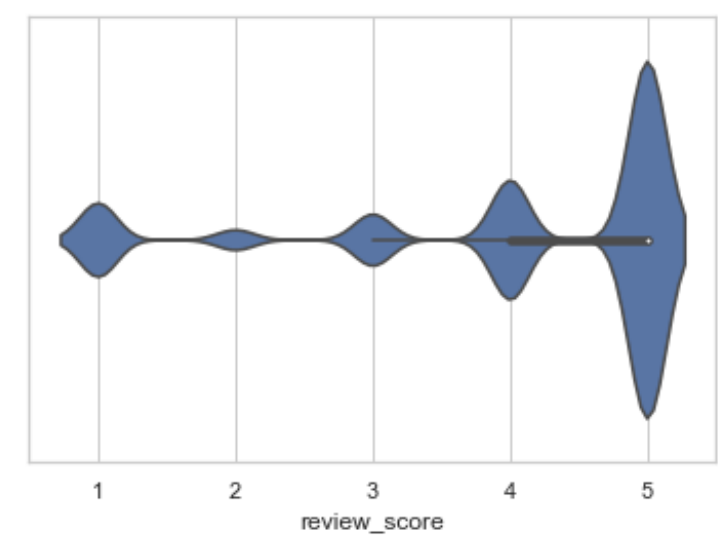
Análise dos Reviews

- Separamos a análise em 2:
 - Clientes com produto entregue sem comentário no review
 - Comentários variados.

Análise dos Reviews

Cientes com produto entregue e sem comentário no review

- A maior parte dos reviews atingiu nota máxima.



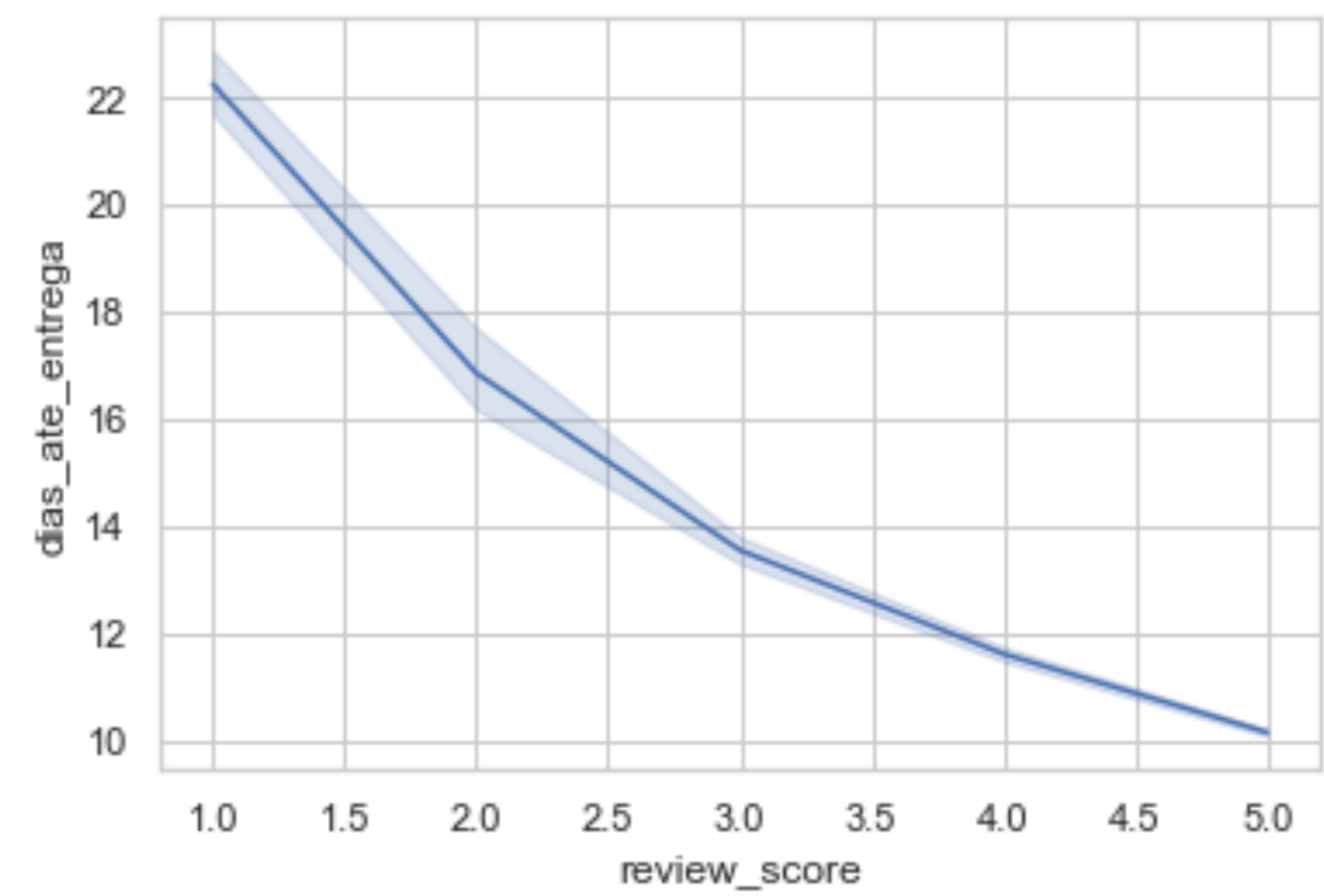
- 76% dos reviews atingiram nota 4 ou superior.

Nota	Percentual
5	57.42%
4	19.2%
1	11.8%
3	8.2%
2	3.2%

- 88% dos reviews está sem título
- 56% dos reviews não tem comentário

- Como seria de se esperar, menos dias até a entrega resultam em score maior.

Relação entre dias até entrega e Score do Review



review_score	dias_ate_entrega
1	22.254663
2	16.850394
3	13.540489
4	11.600768
5	10.135684

- Compras que receberam review com nota máxima levam em média 10 dias para serem entregues, enquanto aquelas que tiveram a pior nota são entregues em quase 30 dias.
- São apenas 3 dias de intervalo na entrega que separam os reviews nota 5 dos nota 3.

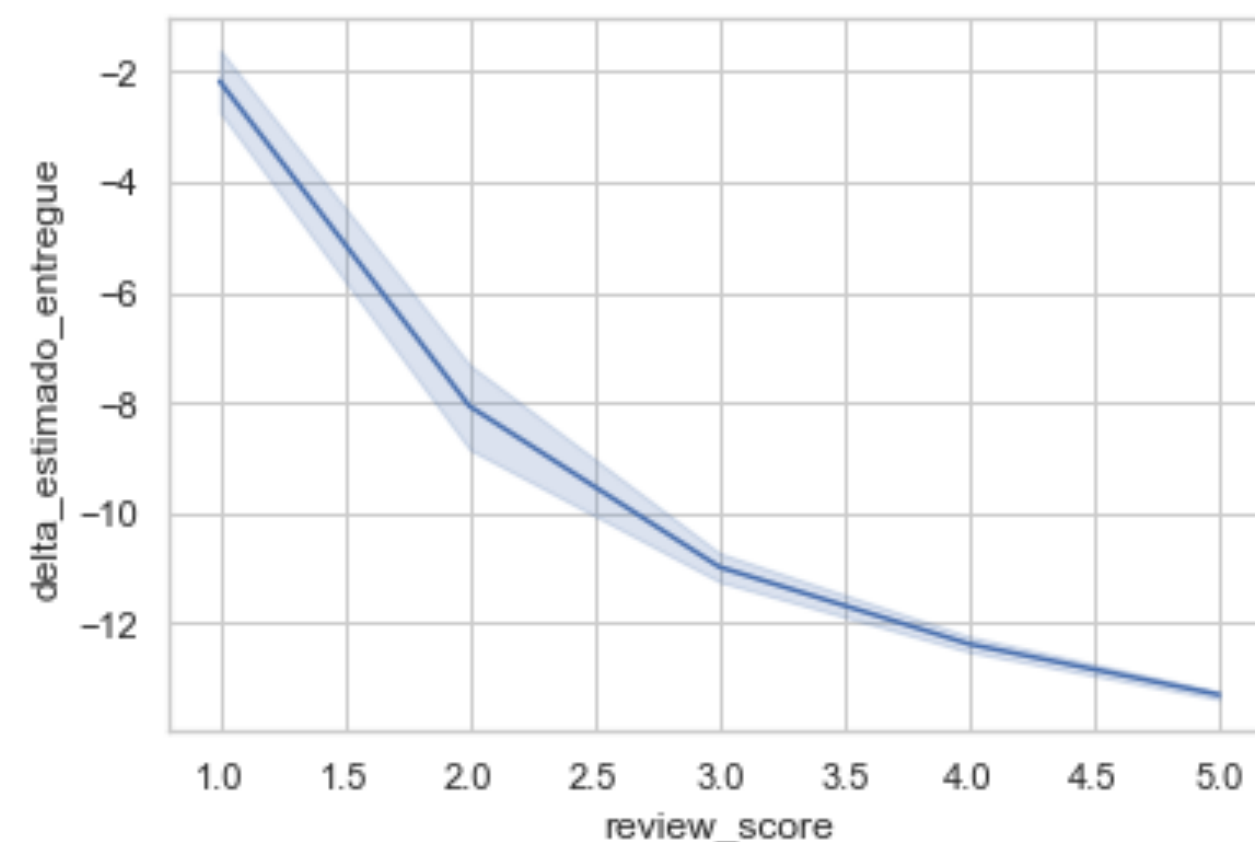
Análise dos Reviews

Clientes com produto entregue e sem comentário no review

- As compras que são recebidas antes da data estimada tendem a receber um review melhor.

no eixo y, quanto mais dias o produto é recebido antes do prazo. maior tende a ser o score do review

Relação entre data estimada, entrega e Score do Review



- O prazo de entrega médio passado para todos os clientes é de 23 dias.

	review_score	delta_estimado_entregue
0	1	-2.167862
1	2	-8.057743
2	3	-10.974599
3	4	-12.371973
4	5	-13.290389

- Podemos perceber acima que encomendas que chegam até 12 dias antes do prazo recebem score 4 ou maior.

■ Conclusão:

- clientes que recebem o produto em até 11 dias depois da compra ficam satisfeitos
 - clientes que recebem o produto em até 11 dias antes do prazo acabar ficam satisfeitos.
-
- Poderíamos lançar uma hipótese e teste A/B para Growth:
 - Aumentar a data de entrega estimada, aumentaria o score dos reviews ? Assim, produtos com reviews mais altos teriam melhor conversão ?

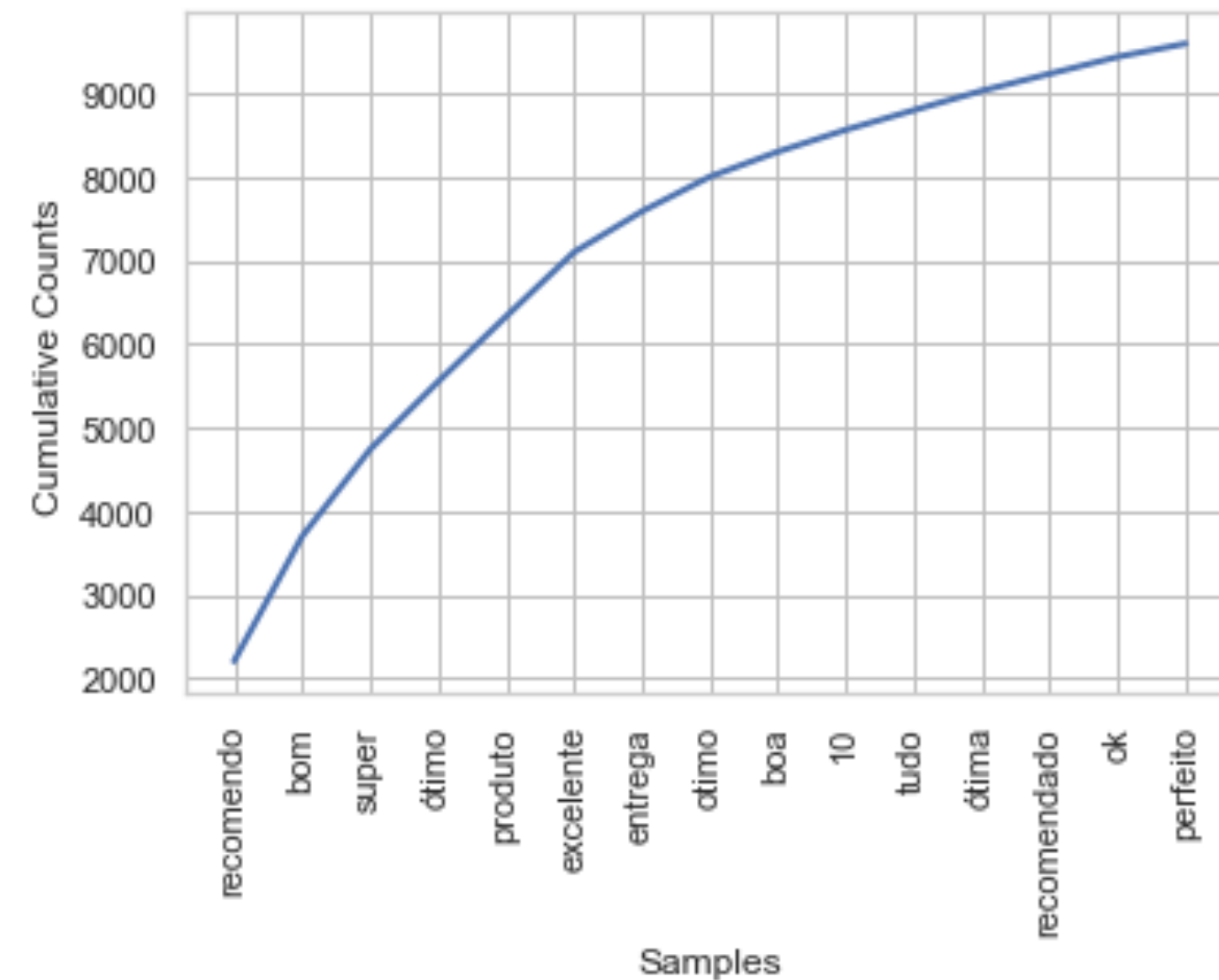
Análise dos Reviews

Análise geral dos comentários

- Separamos a análise textual dos reviews olhando para os títulos e para o conteúdo:
- **+ de 70% dos reviews com título, levam essas 15 palavras:**
‘recomendo’, ‘bom’, ‘super’, ‘ótimo’....

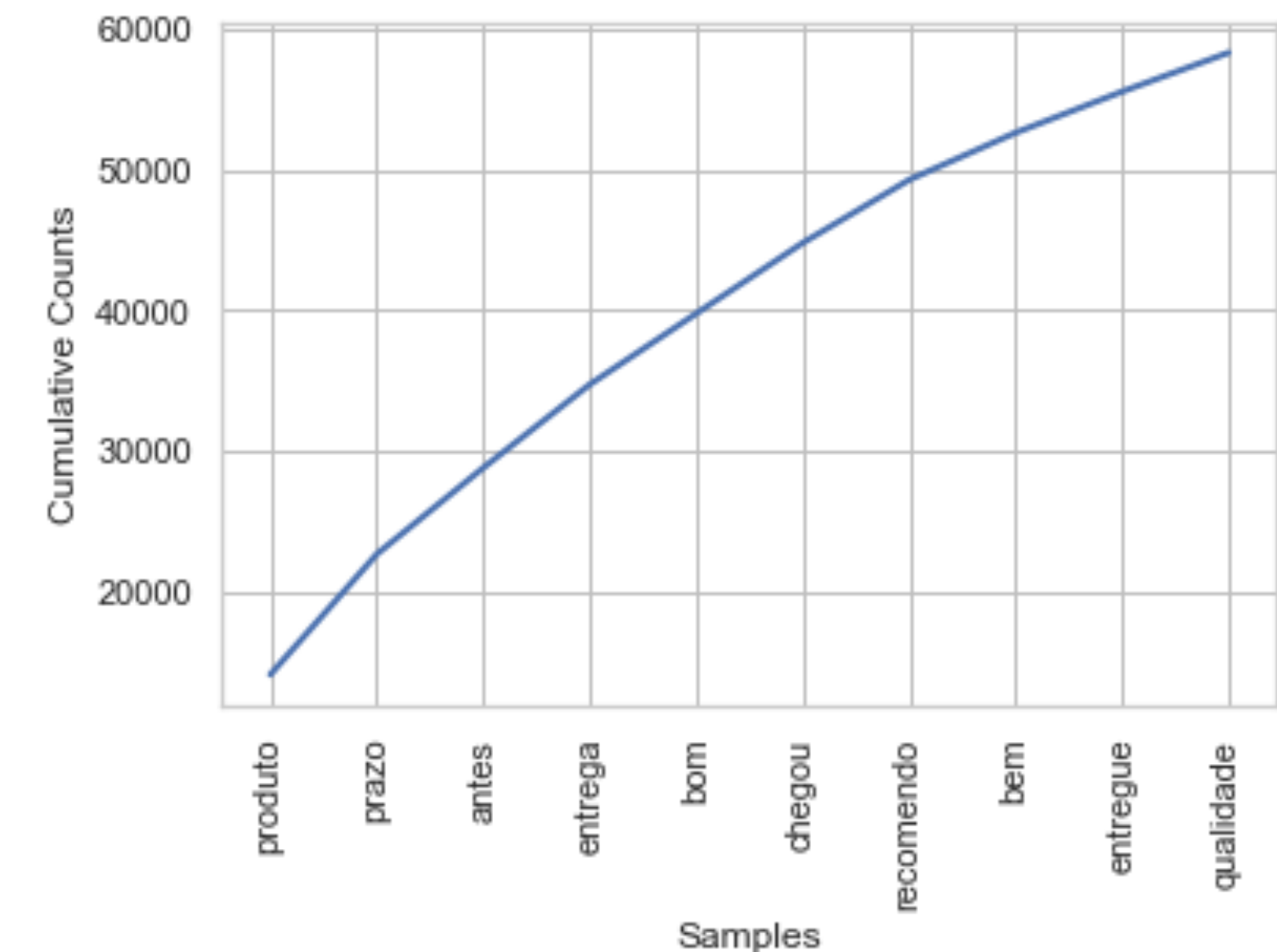
TÍTULOS MAIS COMUNS NOS REVIEWS

Total de Reviews com título: 14189



PALAVRAS MAIS COMUNS NOS REVIEWS

Total de Reviews com comentário: 51252

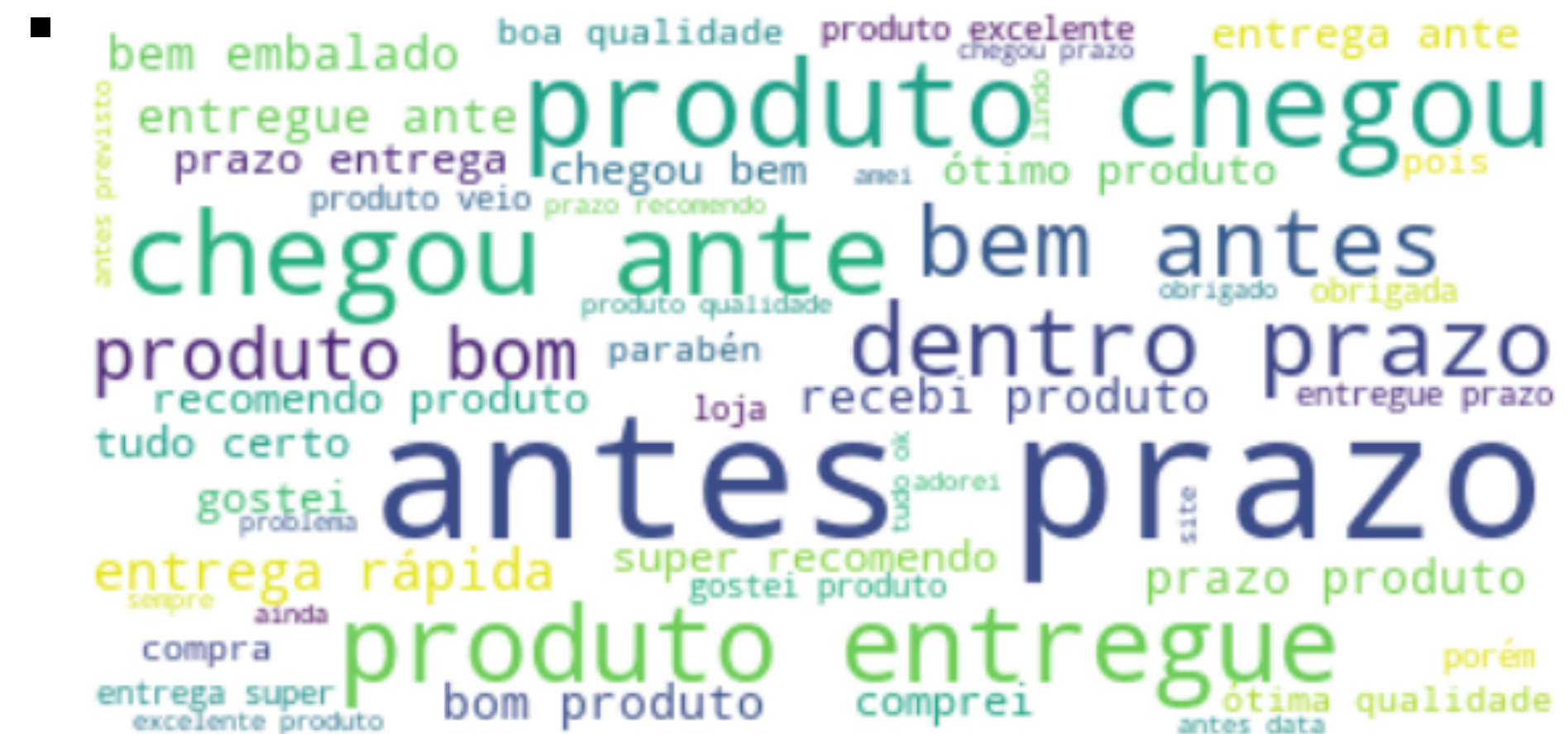


- Estas são as 10 palavras mais citadas no conteúdo dos comentários:
 - ‘prazo’, ‘antes’, ‘bom’, ‘entrega’...

Análise dos Reviews

Análise geral dos comentários

- Wordcloud para melhor percepção dos comentários:



- Fizemos uma análise também dos comentários negativos (Clientes insatisfeitos somaram 15% das análises)

- **As principais causas se referiram à:**

- Pedido não entregue no prazo..
- Recebimento do pedido de forma parcial.
 - Esta análise está no notebook

- Foi criado um algoritmo para classificar os reviews em: Review Positivo e Review Negativo. A apresentação deste está no final.

Análise de Produtos

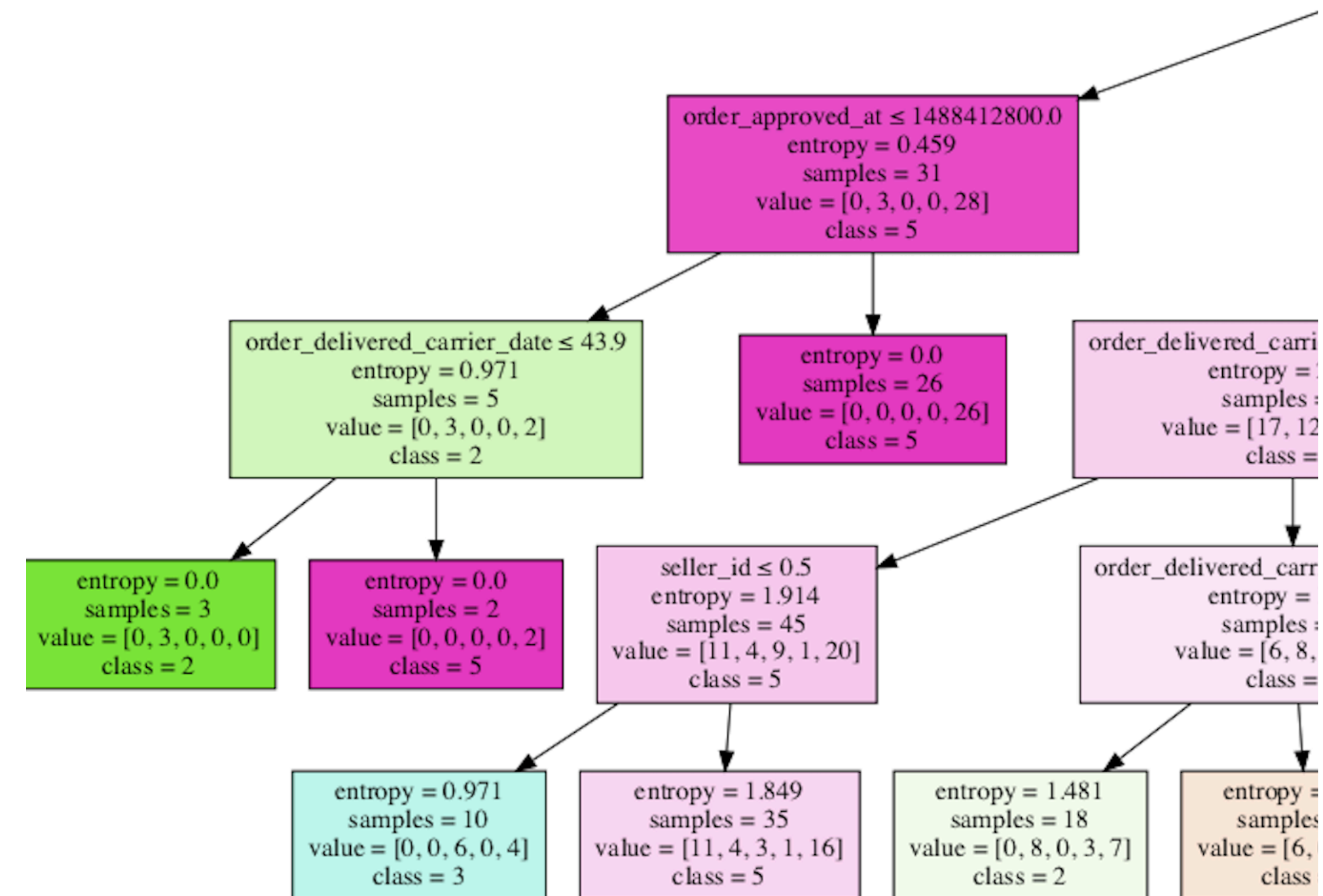
- Foram separadas as vendas com score abaixo de 3
 - As vendas foram agrupadas por categoria e por estado
 - Foram contadas as categorias de produto por estado com Score baixo
-
- **Categorias mais propensas a insatisfação:**
 - **Cama Mesa & Banho**
 - **Acessórios de Informática**
 - **Móveis e Decoração**
 - **SP, RJ e MG são os estados com mais reclamações**
-
- **Sabemos que Móveis e acessórios são itens grandes, e que entrega atrasada gera review baixo.**
 - **Como experimento, o sistema poderia dar preferência na exibição para o cliente, de lojistas próximos à sua cidade.**



Prevenendo a Insatisfação

Modelo de árvore de decisão

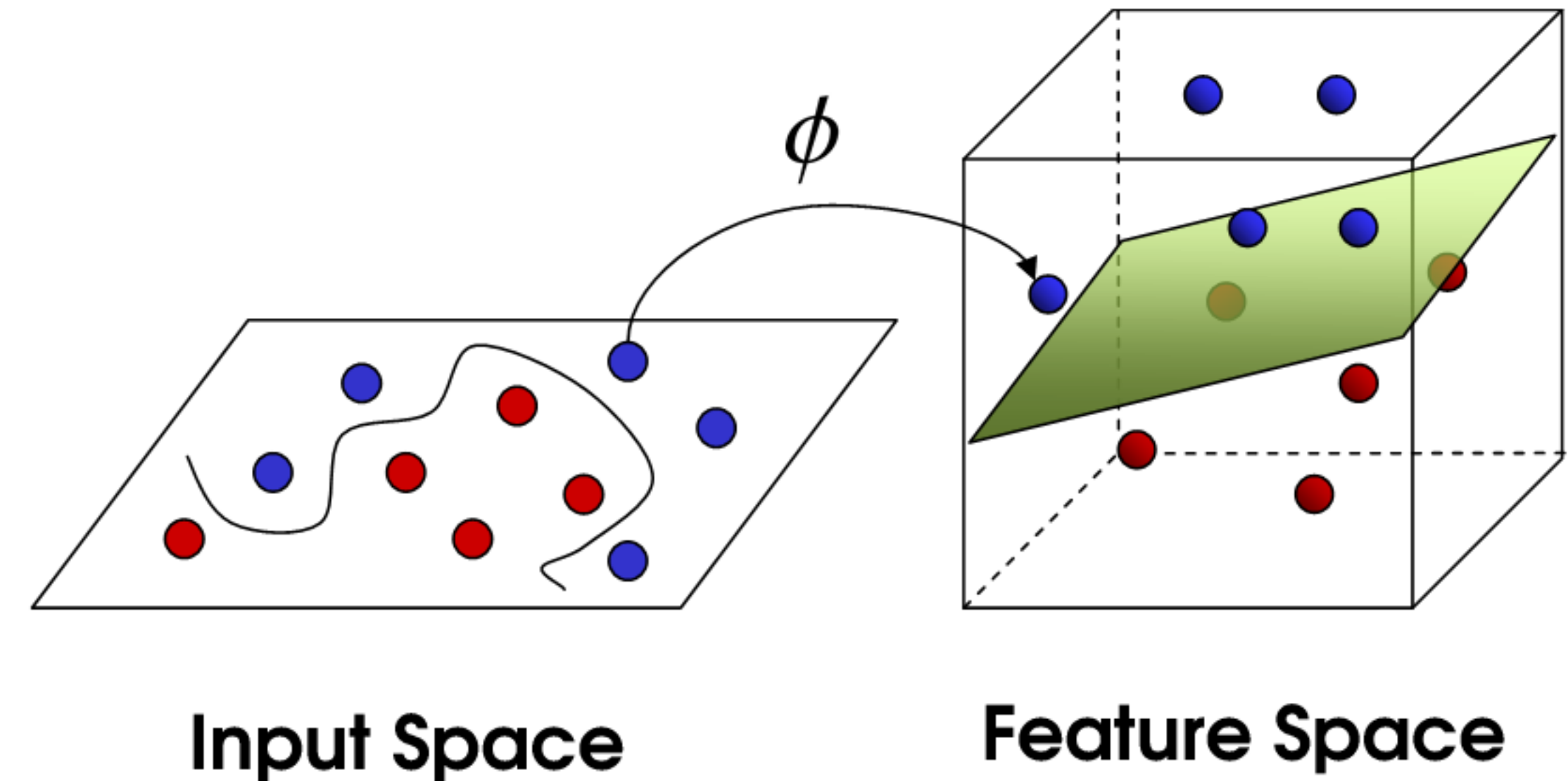
- Buscando ter como base possíveis futuros desafios no time de Marketing, criamos um algoritmo para prever clientes insatisfeitos.
- Foram usadas 14 das features fornecidas no teste
- As features tem dados da compra do cliente, cidade, infos do produto, vendedor e etc.
- Entre testes e ajustes o modelo atingiu uma acurácia de 58% na previsão do Score das compras efetuadas
 - Em um teste o modelo previu 746 possíveis reviews negativos (considere aqueles com nota menor que 3)
- **Este Algoritmo pode ajudar o time a se antecipar perante possíveis detratores e criar gatilhos que surpreendam esses clientes com uma experiência positiva e que os desencorajem a postar reviews negativos**



“Medindo a Febre”

Modelo de análise de sentimento com SVM

- Para complementar o projeto envolvendo prever clientes insatisfeitos antes mesmo de receber o produto
- Desenvolvemos também um modelo para classificar os reviews
- Foi utilizada a técnica Support Vector Machine
- Entre testes e ajustes o modelo atingiu uma acurácia de 92% na classificação de reviews positivos e 91% para reviews negativos
- Este Algoritmo pode ajudar o time a mais rapidamente classificar comentários negativos e tomar ações corretivas.



- **Categorias mais propensas a insatisfação:**
 - Cama Mesa & Banho
 - Acessórios de Informática
 - Moveis e Decoração
- **Clientes que recebem o produto em até 11 dias antes do prazo acabar ficam satisfeitos.**
- **Como experimento, o sistema poderia dar preferência na exibição para o cliente, de lojistas próximos à sua cidade.**
- **SP, RJ e MG são os estados com mais reclamações**
- **Clientes que recebem o produto em até 11 dias depois da compra ficam satisfeitos.**
- **Aumentar a data de entrega estimada, aumentaria o score dos reviews ? Assim, produtos com reviews mais altos teriam melhor conversão ?**
- **Algoritmo para prever os Insatisfeitos e Classificar os Reviews**