NEGATIVE CUSTOMER REVIEWS

PREDICTIVE CLASSIFICATION MODEL

PREPARED FOR OLIST

JUNE 17, 2021

Flatiron Data Science Bootcamp Project

insight CONSULTING

TODAY'S MODEL REVIEW

Outline

- Business Objective
- Visual Highlights
- Bad Review Model Solution
- Model Feature Importance
- Actionable Next Steps

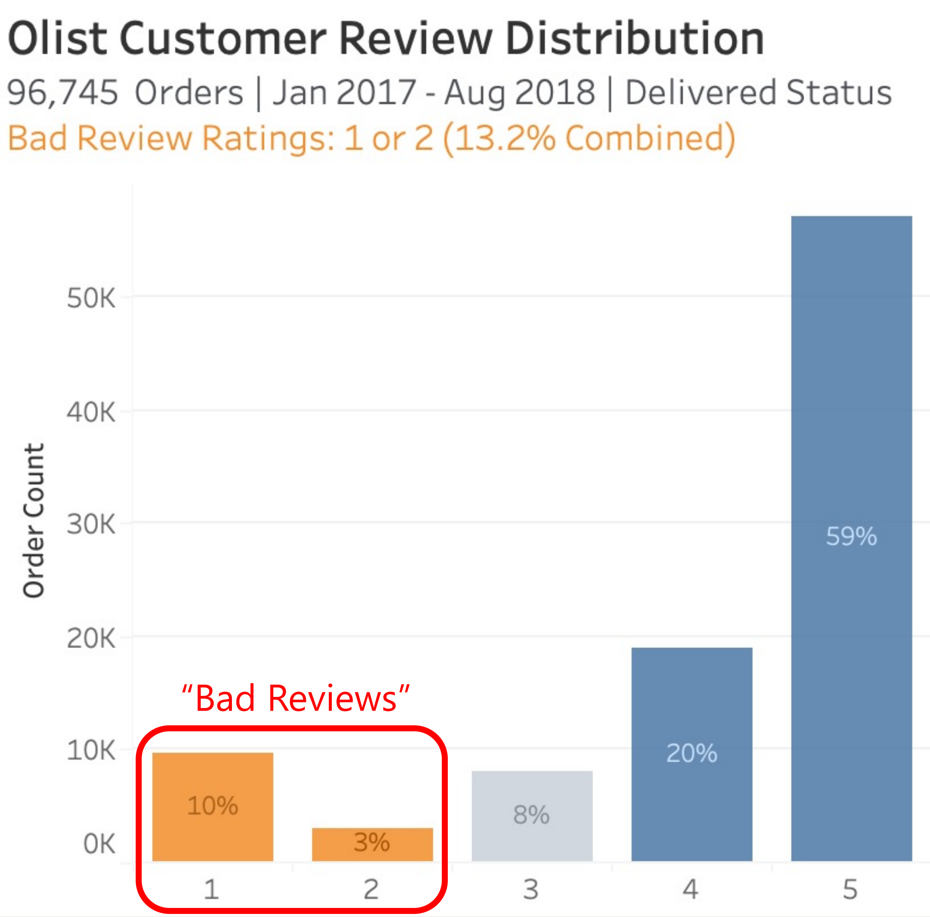
PREDICTING BAD CUSTOMER REVIEWS

Business Objective

Business Objective: Understand and predict bad customer reviews (scores 1 or 2) in order to try to minimize bad reviews and improve customer satisfaction.

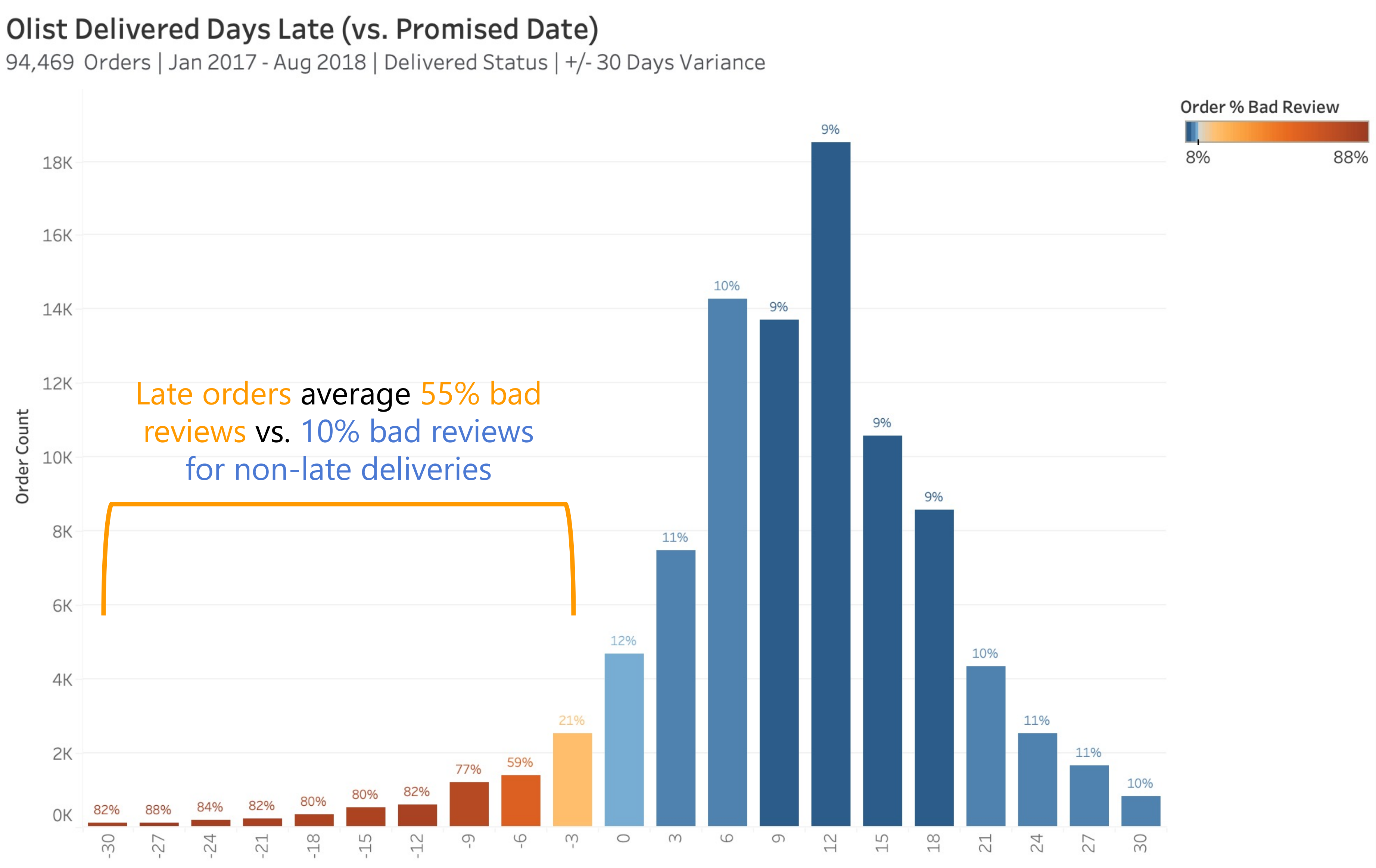
Requirements

- Understand dynamics that cause customers to leave a bad review
- Find the best prediction model to classify a bad review from a customer
- Deliverables should explain the relative influence of each predictor on the model
- Deliverables should suggest potential solutions to preventing bad customer reviews



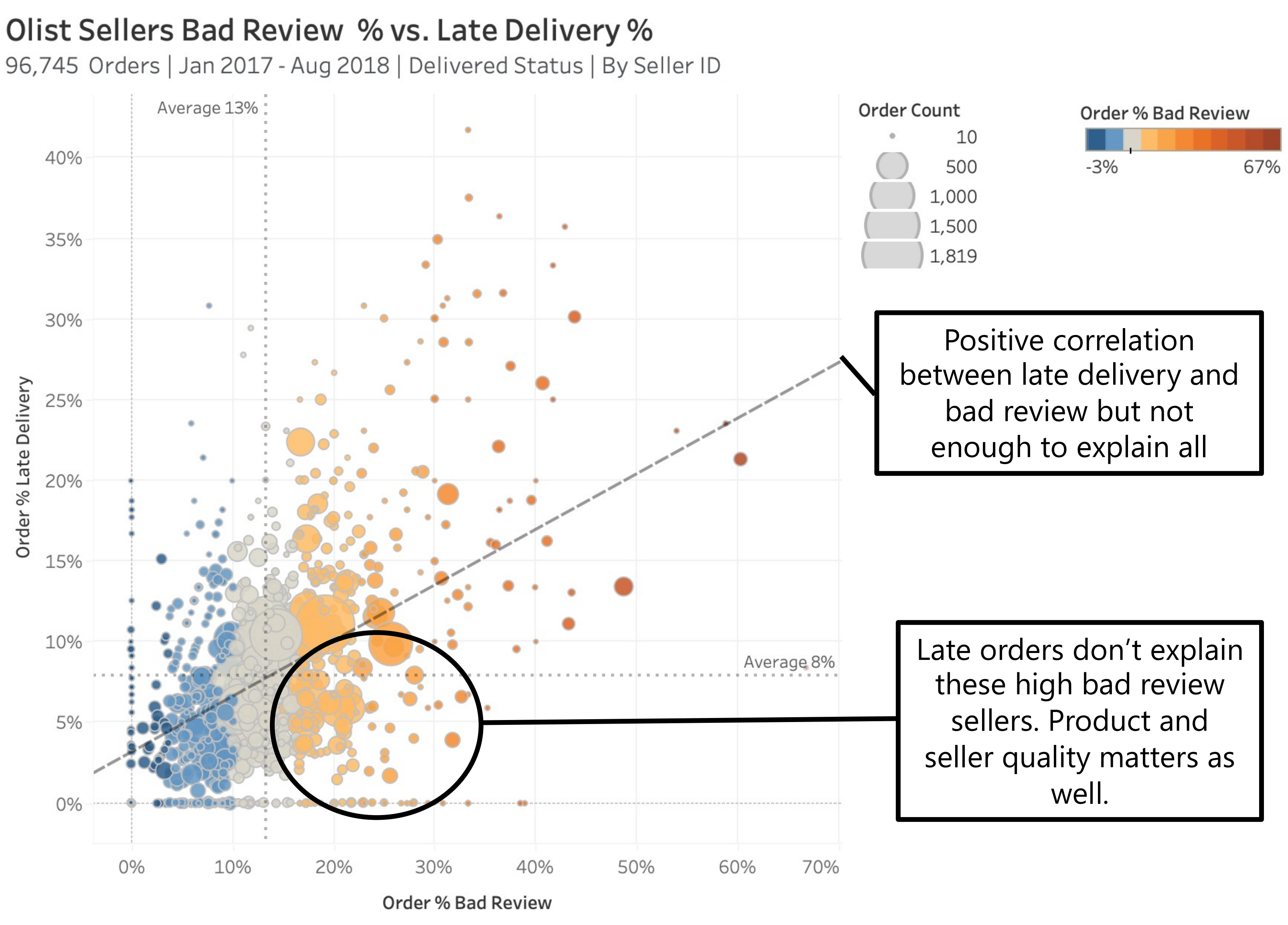
MEETING PROMISED DELIVERY DATES

Late Delivery



SELLER QUALITY MATTERS

Seller Quality



Positive correlation between late delivery and bad review but not enough to explain all

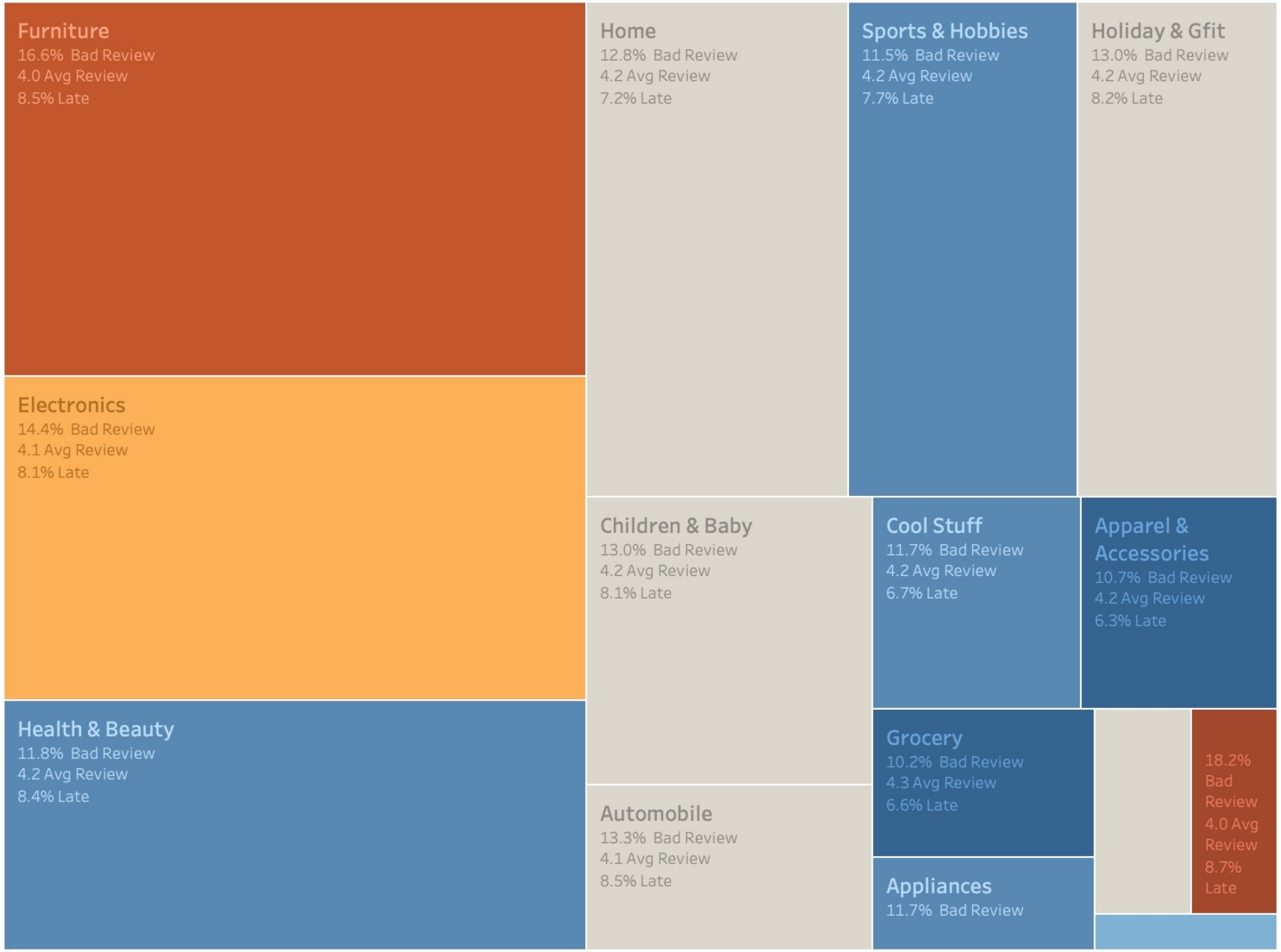
Late orders don't explain these high bad review sellers. Product and seller quality matters as well.

MARKETPLACE PRODUCT VARIETY

Product Category

Olist Product Category Distribution

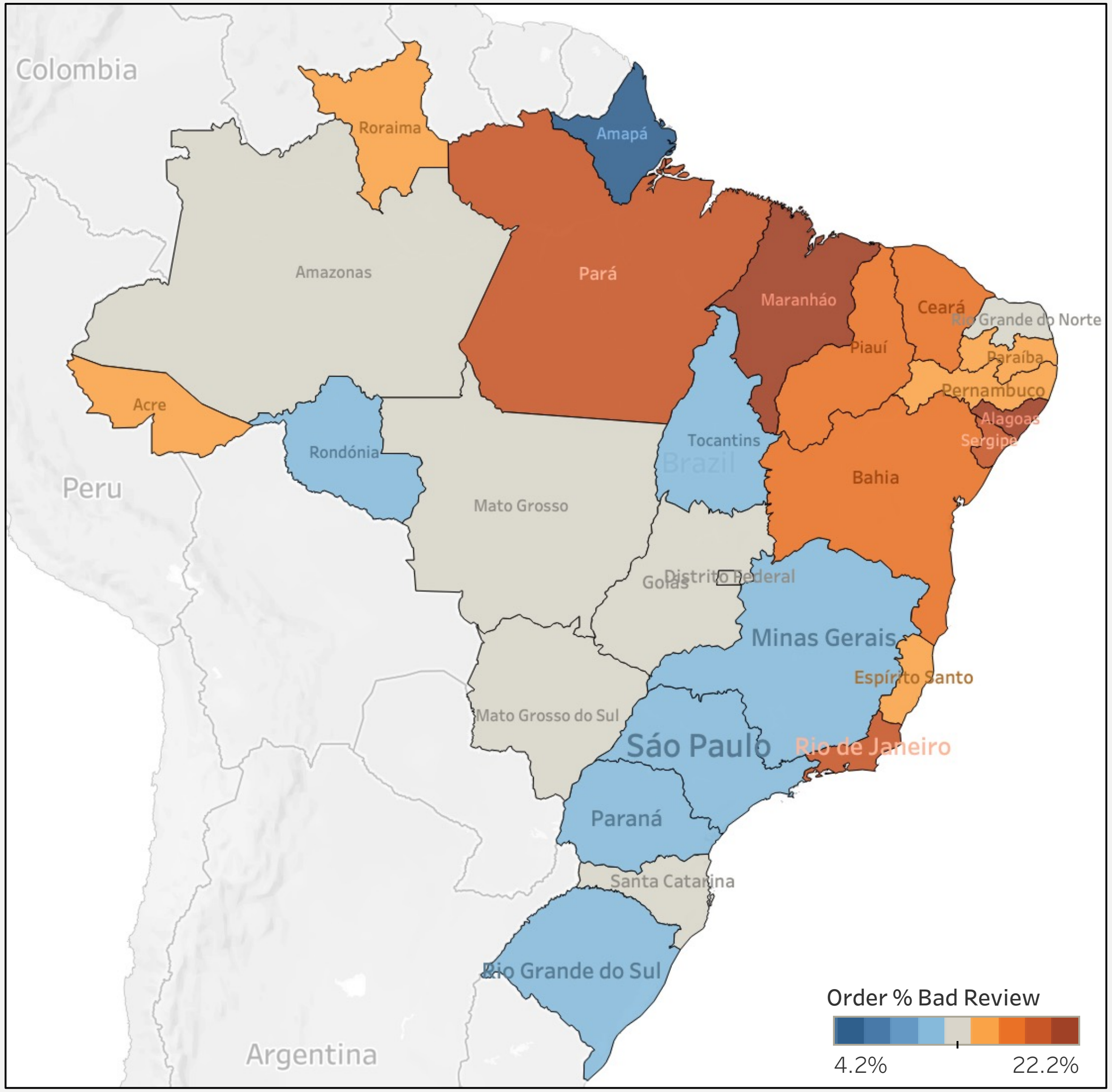
96,745 Orders | Jan 2017 - Aug 2018 | Delivered Status



The bad review rate (%) for each unique product varies substantially and many aren't meeting customer expectations

Product Id	Product Category Parent	Order % Bad Review	Order Count
fd0065af7f09af4b82a0ca8f3eed1852	Automobile	90.0%	10.0
6d2fde7d12bb6ff367dbda120ba8828e	Electronics	83.3%	12.0
cd46a885543f0e169a49f1eb25c04e43	Electronics	82.1%	28.0
b36f3c918c91478c4559160022d3f14e	Other	73.3%	15.0
b1d207586fca400a2370d50a9ba1da98	Other	71.4%	42.0
16bf176650a888512655cc94f61860e3	Holiday & Gfit	63.6%	11.0
89b121bee266dcd25688a1ba72eefb61	Electronics	59.6%	57.0
86ecc269de40ba13205e7beeee12f26f	Electronics	58.3%	12.0
25c38557cf793876c5abdd5931f922db	Children & Baby	55.3%	38.0
7b35ccd93a2184646c03b70326626923	Sports & Hobbies	55.0%	20.0
43ee88561093499d9e571d4db5f20b79	Furniture	53.8%	13.0
63085bb4366ded27bcb63cbb59b4103a	Children & Baby	53.8%	13.0
0cf41187284d7f099adc8415a743ebbd	Children & Baby	53.1%	32.0
fb01a5fc09b9b9563c2ee41a22f07d54	Electronics	52.2%	23.0
1dec4c88c685d5a07bf01dcb0f8bf9f8	Automobile	51.4%	35.0

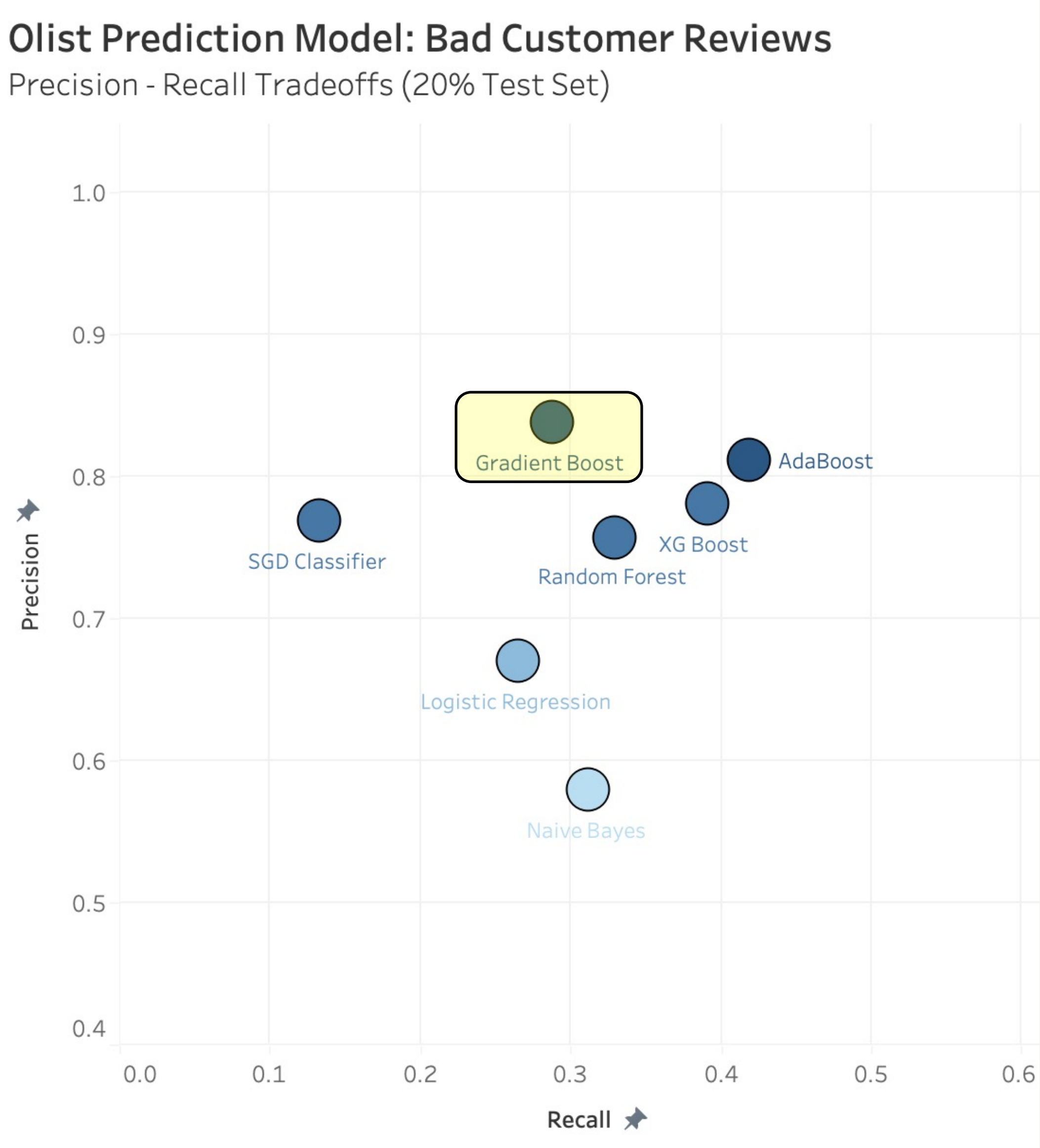
REVIEWS VARY BY BRAZIL STATE



Customer State	Order % Bad Review	Delivery Days	Order % Late	Product Index
Acre	15.0%	21	3.8%	122
Alagoas	22.2%	25	23.7%	164
Amapá	6.0%	27	3.0%	49
Amazonas	13.8%	26	4.1%	110
Bahia	17.7%	19	13.8%	140
Ceará	17.7%	21	15.3%	137
Distrito Federal	13.5%	13	7.0%	109
Espírito Santo	14.7%	16	12.2%	117
Goiás	13.8%	16	8.0%	112
Maranháo	20.5%	22	19.2%	165
Mato Grosso	13.2%	18	6.6%	118
Mato Grosso do Sul	13.4%	16	11.1%	108
Minas Gerais	12.1%	12	5.5%	98
Pará	18.8%	24	12.4%	149
Paraíba	15.3%	20	11.0%	121
Paraná	11.2%	12	4.8%	95
Pernambuco	15.2%	18	10.7%	122
Piauí	16.6%	19	15.6%	134
Rio de Janeiro	18.9%	15	13.3%	148
Rio Grande do Norte	13.6%	19	10.9%	117
Rio Grande do Sul	12.1%	15	6.9%	104
Rondónia	11.9%	19	2.9%	103
Roraima	15.0%	30	12.5%	120
Santa Catarina	13.5%	15	9.6%	110
Sáo Paulo	11.0%	9	5.7%	92
Sergipe	19.3%	21	15.4%	148
Tocantins	12.0%	18	12.4%	90
Grand Total	13.2%	13	7.9%	111

MODEL PERFORMANCE COMPARISON

Model Summary

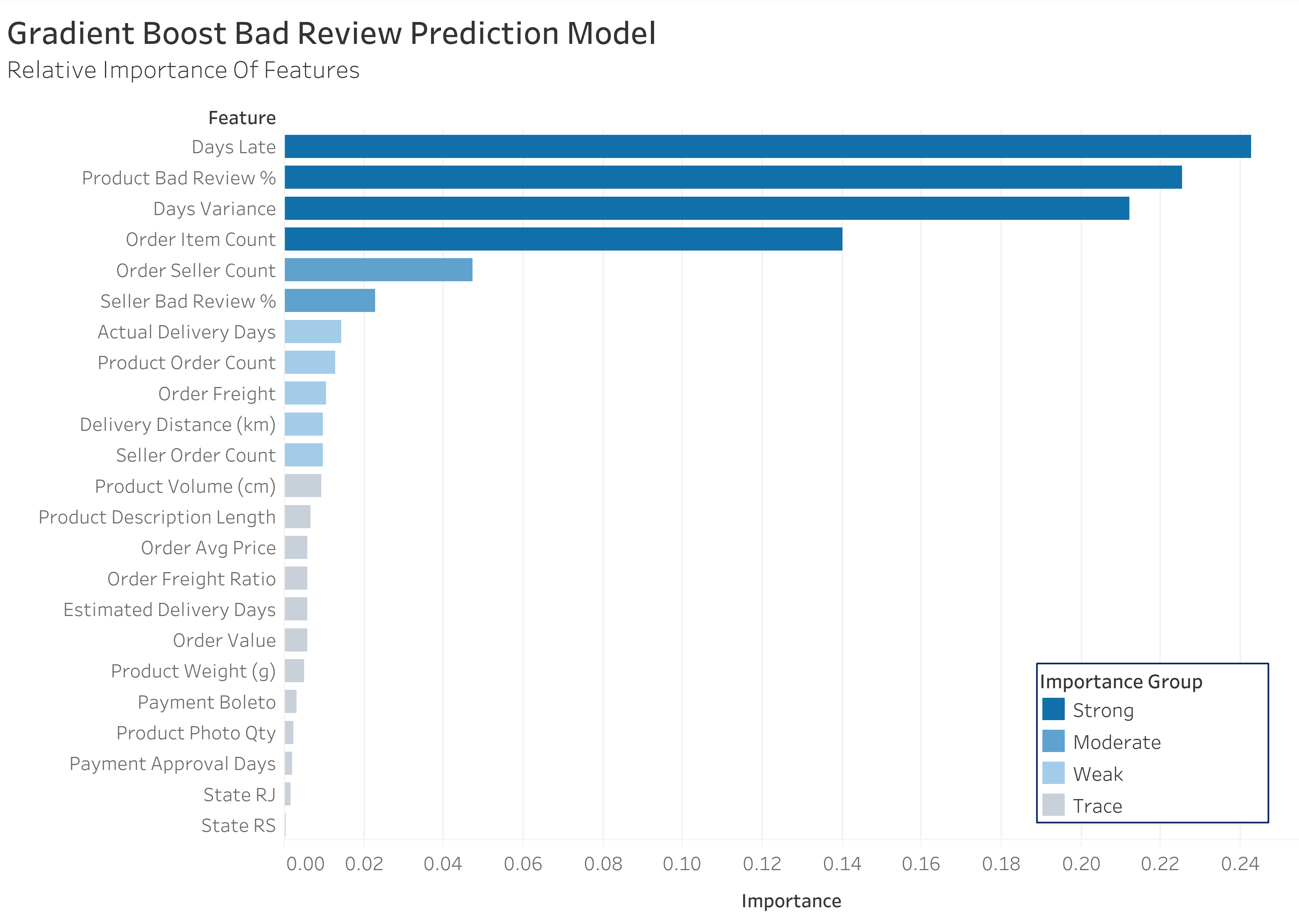


Recommend Gradient Boost Model

- Predicts bad reviews with 84% precision
- Captures 29% of bad reviews (recall)
- Even higher precision possible for lower recall (90% precision @ 12% recall)
- Understandable feature importance
- Model performs relatively efficiently

MODEL PREDICTOR IMPORTANCE

Gradient Boost Model



Strong Impact

- Delivered days late
- Days variance (vs. estimated)
- Products avg. bad review rates
- Count of items on order

Moderate Impact

- Count of sellers on order
- Sellers avg. bad review rates

Weak Impact

- Actual delivery days
- Total orders for product
- Total orders for seller
- Freight on order
- Delivery distance in kilometers

NEXT STEPS

1. Define customer retention actions
2. Perform randomized trials on treatment/control groups
3. Measure ROI uplift from each action/experiment
4. Formalize high-performing actions into business

Predictive capabilities may be expanded with broader customer data profiles. Explore potential data such as:

- Review Comment – Parse text sentiment to derive insight into customer complaints
- Customer Service – Incorporate Contacts/Call/Emails, Tickets, Website/App Usage
- Demographics – Credit Score, Income, Home Ownership, Household Size, Persona/Lifecyle, Time at Address, Other customer demographics
- Supplier – Incorporate supplier product, process, inventory and quality metrics

"In God we trust, all others bring data." – Edward Deming

Insight Consulting Group is a boutique consulting firm specializing in analytics, business intelligence, strategy, data visualization, and data warehousing for Fortune 500 client accounts.

Visit www.insightconsultinggrp.com for more information.

Please contact us at:

Chuck Utterback, Principal 770.298.3169
chuck.utterback@insightconsultinggrp.com

https://github.com/cutterback
https://www.linkedin.com/in/chuckutterback/

THANKS