Does Lannister always pay his debts?

The OLIST story

olist store

Leveraging historic data for customer satisfaction insights

Petya Georgieva, Violeta Kotova , Maria Katzarova

Telerik Academy Upskill: Data Analyst – Final Project Presentation

Virtual, Sofia

Telerik Academy

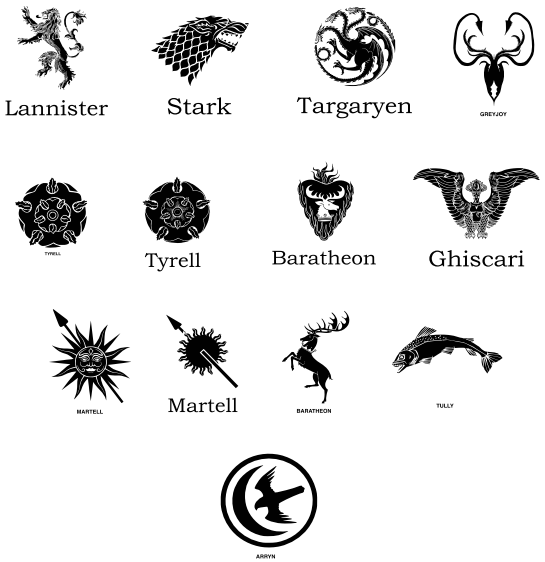
Data overview

Customers and Sellers

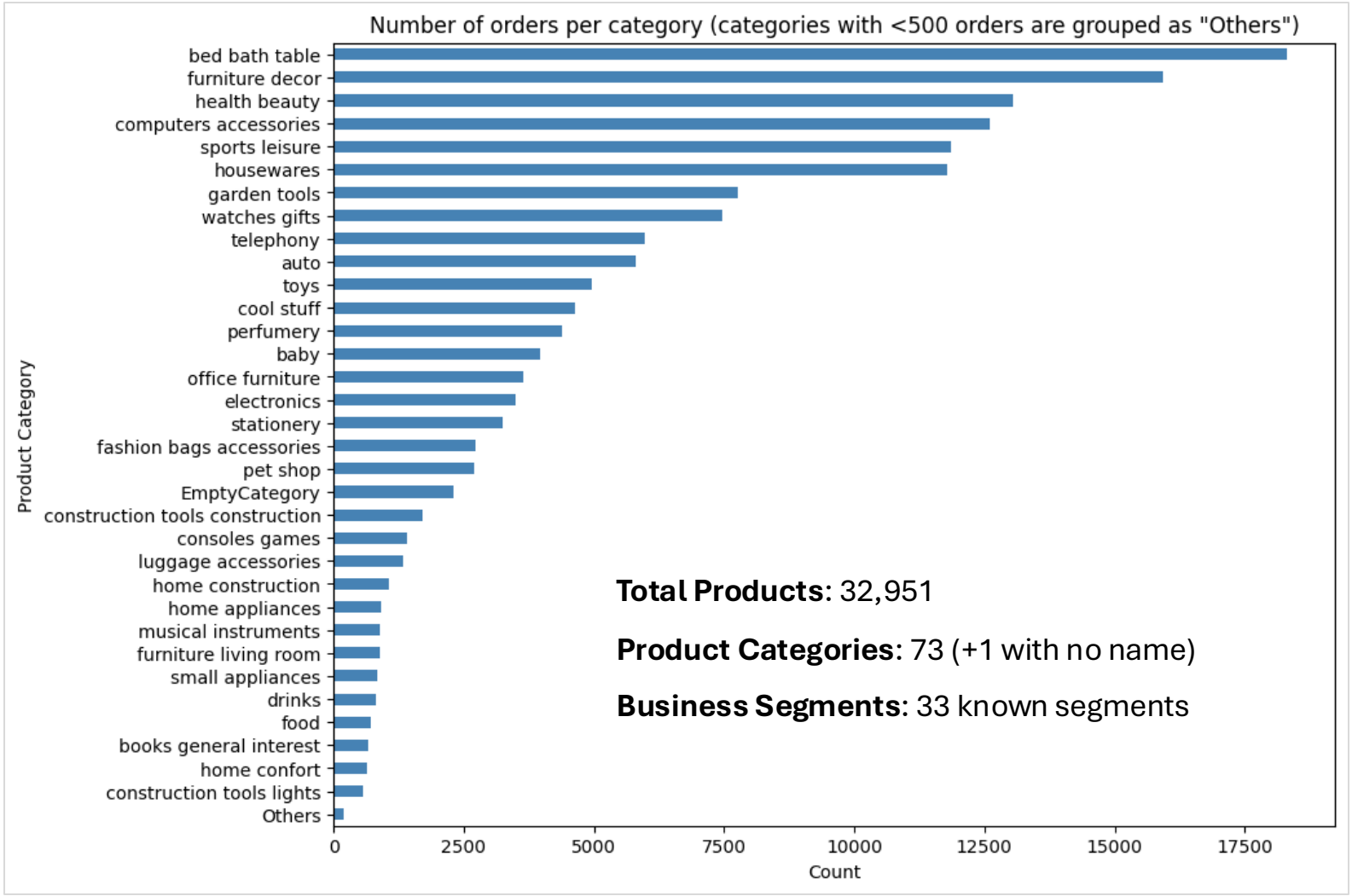
Unique Customers: 96,096

- **Customers with >1 orders**: 2,997
- **Maximum orders by a single customer**: 17

Unique Sellers: 3,095



Products



Total Products: 32,951

Product Categories: 73 (+1 with no name)

Business Segments: 33 known segments

Orders

Total Orders: 99,441

- **delivered orders**: 96,478 (~97%)
- **orders with 1+ items**: 3,236

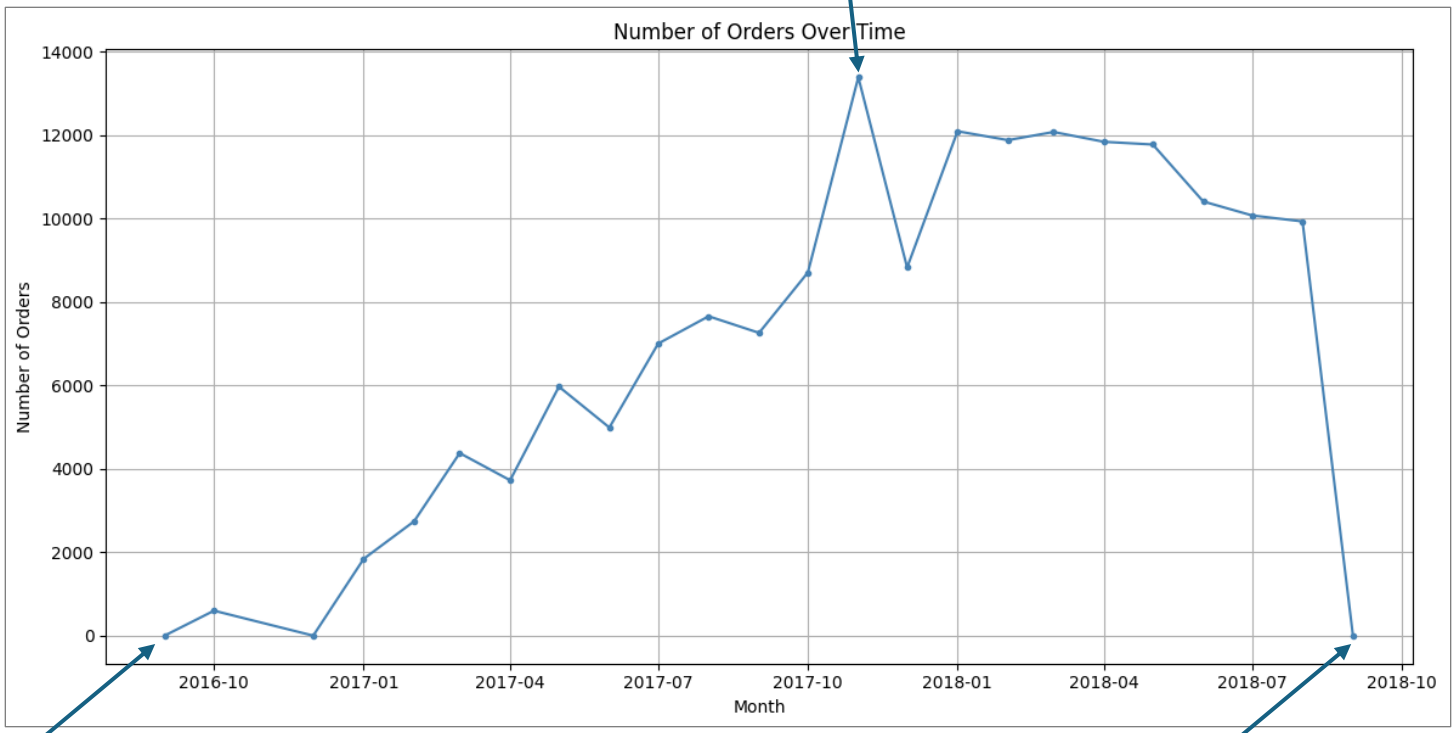
Cost
75% of Orders:

- Cost < R$ 150 (~$27)
- Shipping < R$ 24 (~$4)

Orders Paid with Multiple Payment Forms: 3%
(e.g., credit card, cash, voucher)

Time Frame

Peak of purchases: 13,380 - September 2017

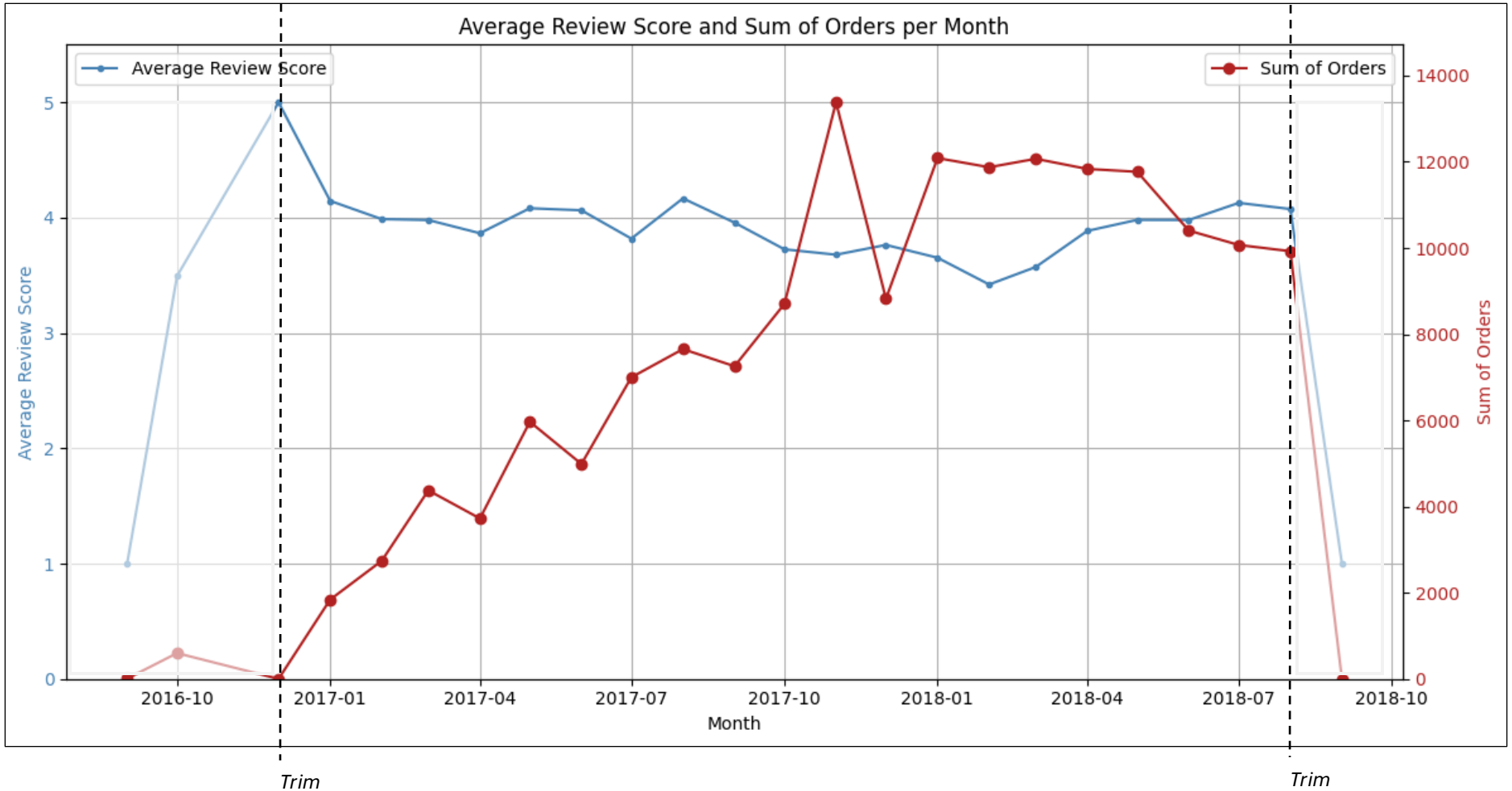


First purchase: September 4, 2016

Last purchase: September 3, 2018

Overview

Customer satisfaction



Total Review Scores: 99,224
out of 99,441 orders

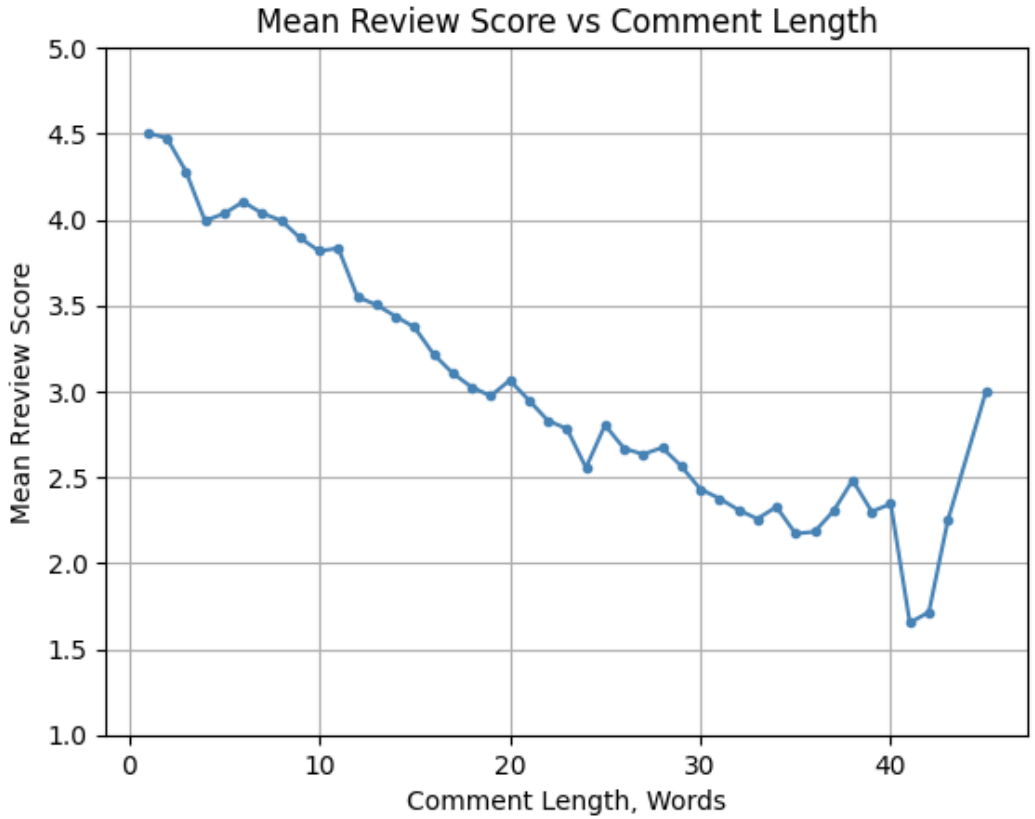
1(Worst) - 5 (Best)

What drives customer satisfaction?

Let's investigate!

Do the customers have something to say?

Comment length: 41% of the reviews have comments

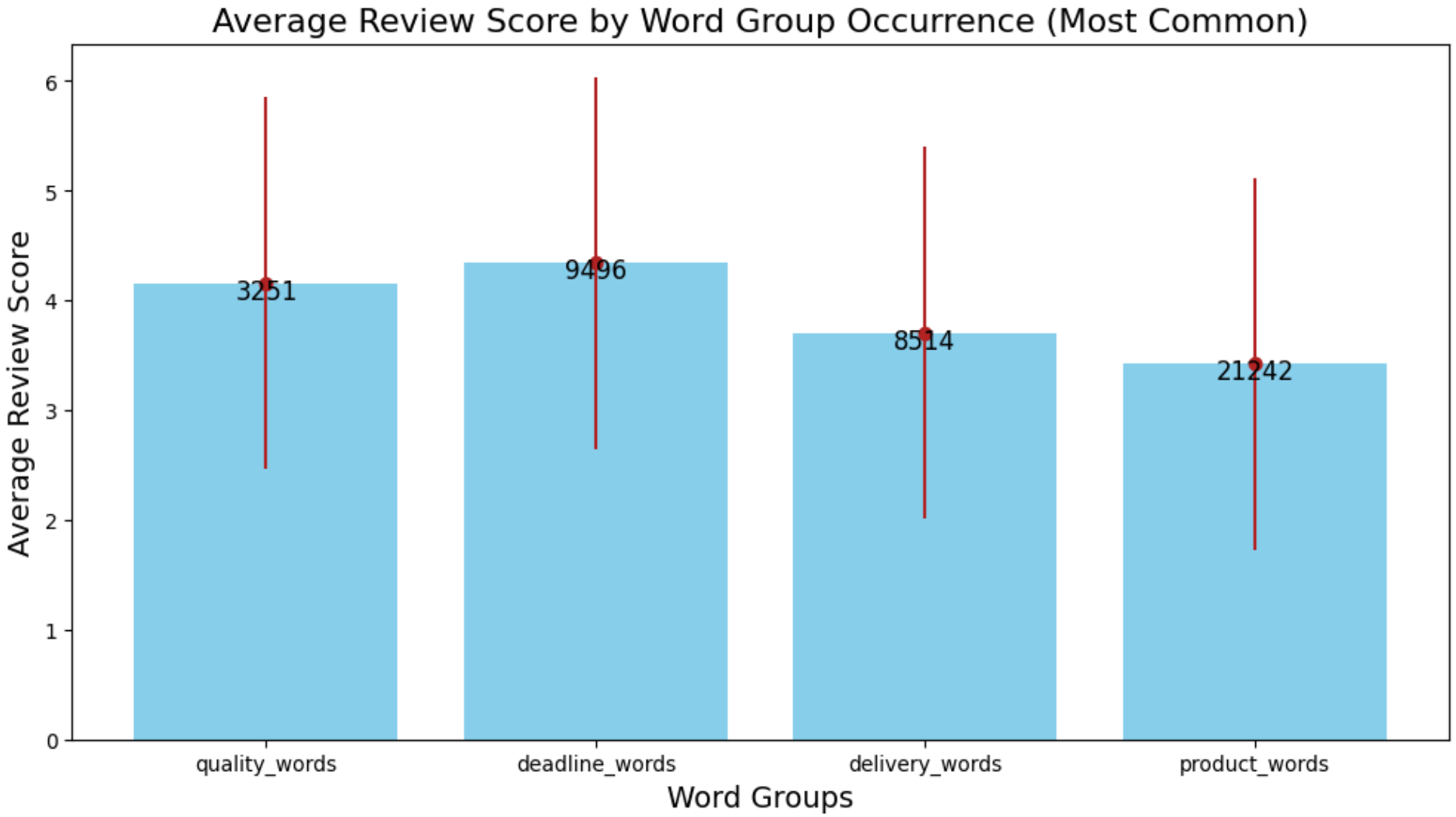


If they are ***unhappy,***
they have a lot to say!

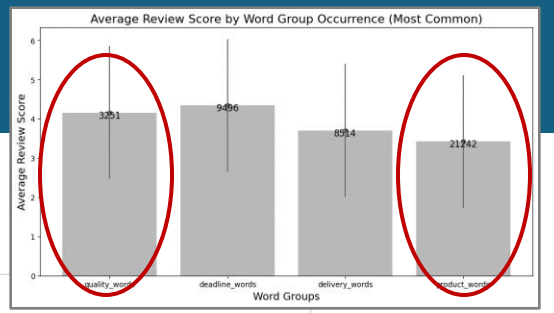
What are the customers saying?

The customer is ***always*** right!

Most common words



Product | Quality



Mean Review Score and Mean Delivery Time by Product Category (>1000 orders)

Mean Review Score

Mean Delivery Time

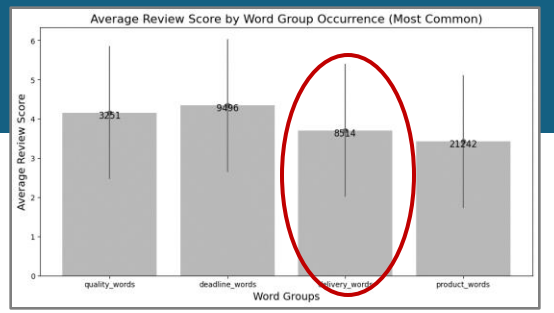
Mean Review Score

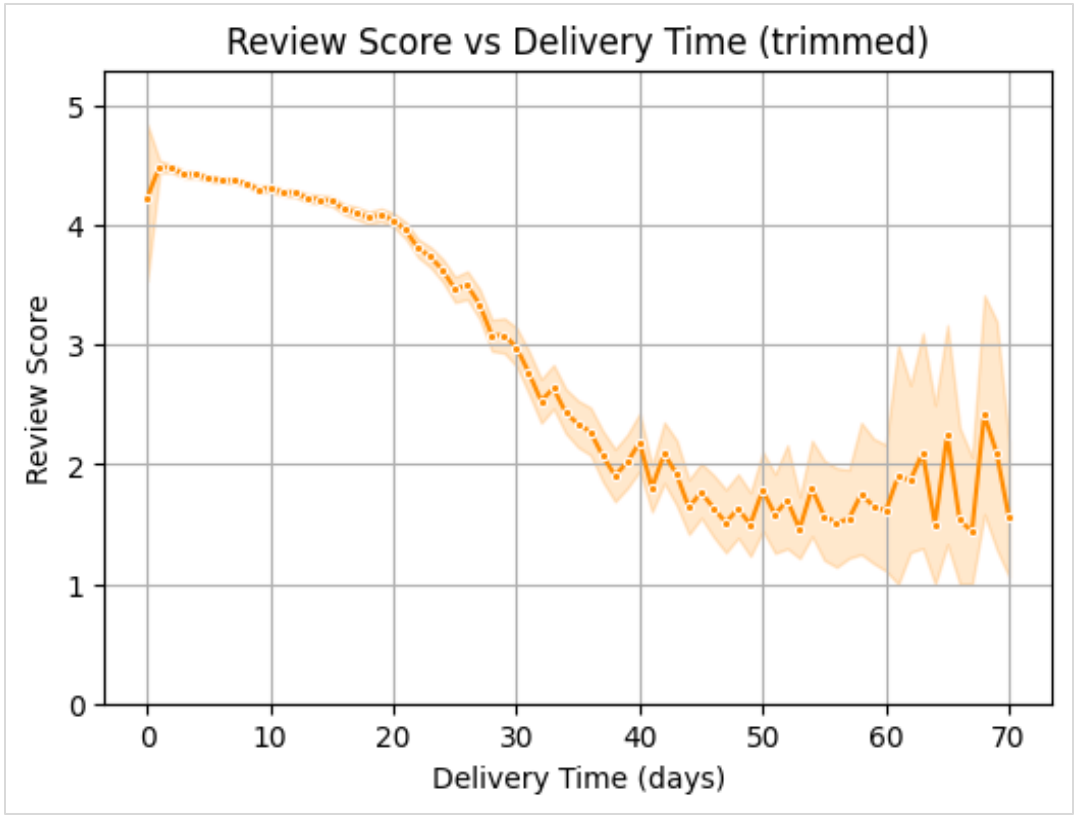
Mean Delivery Time (days)

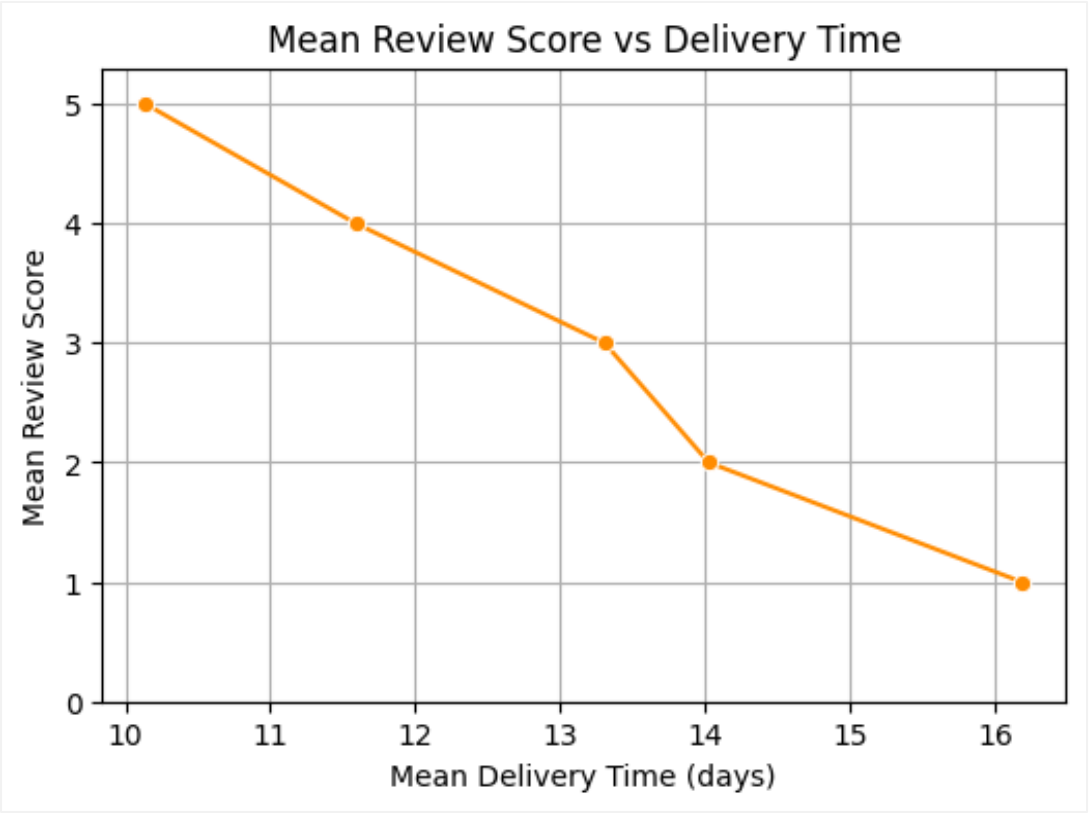
Product Category

luggage accessories
pet shop
toys
fashion bags accessories
stationery
cool stuff
perfumery
auto
construction tools construction
health beauty
sports leisure
electronics
consoles games
baby
watches gifts
home construction
housewares
garden tools
telephony
furniture decor
bed bath table
computers accessories
EmptyCategory
office furniture

Delivery | Deadline

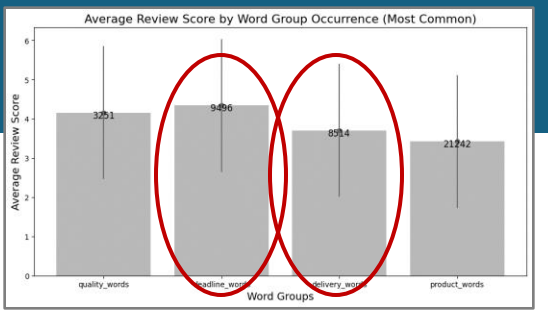


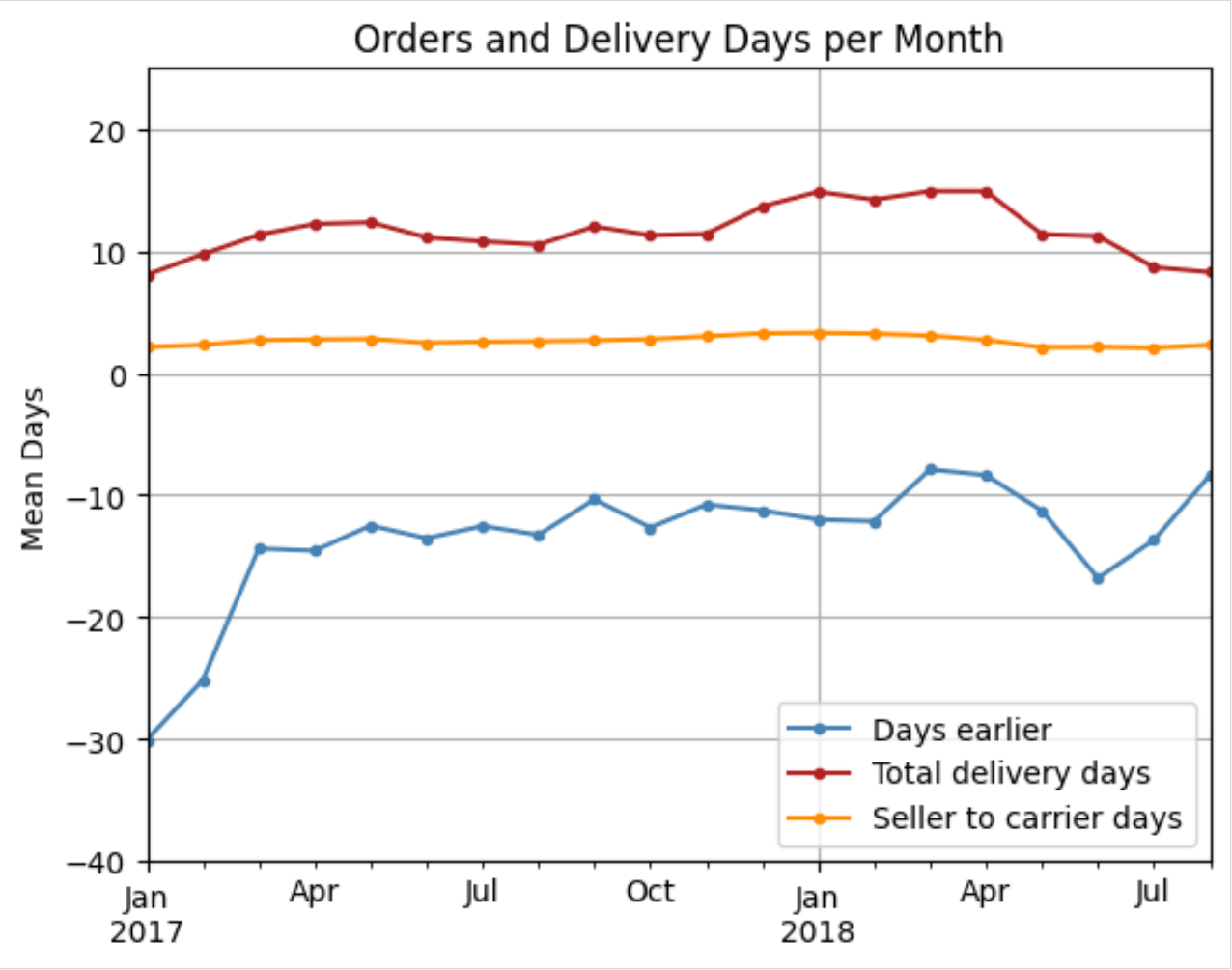




157 orders with delivery days > 70 are trimmed

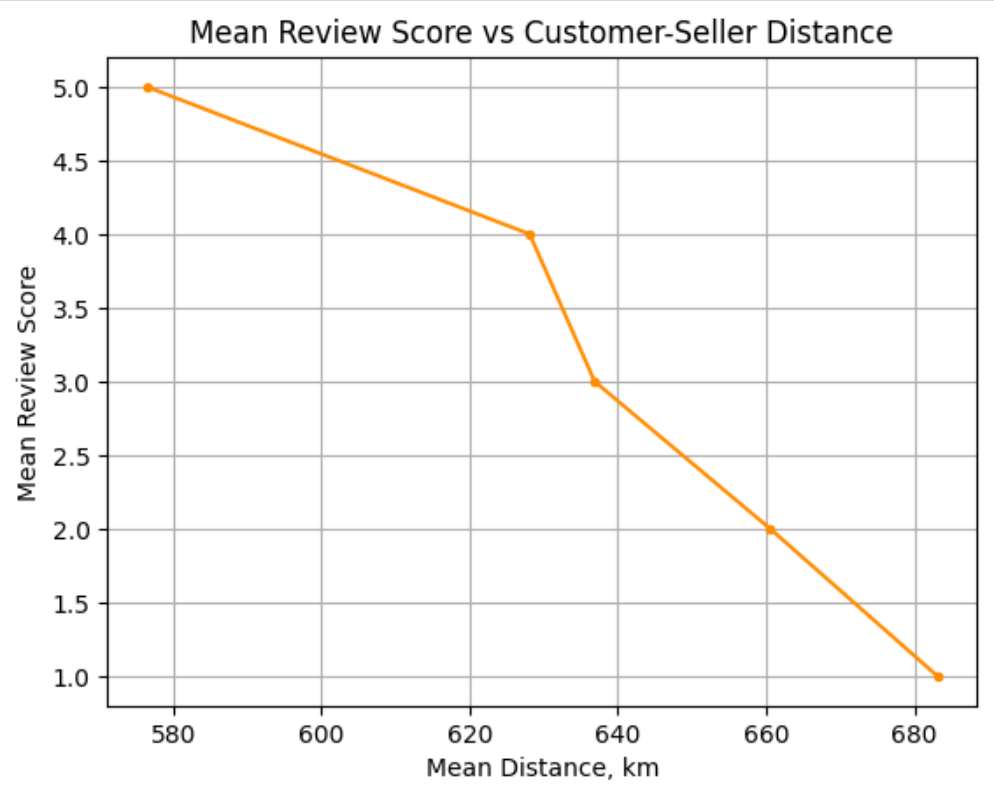
Delivery | Deadline





It is not about the deadline but rather the ***waiting time***!

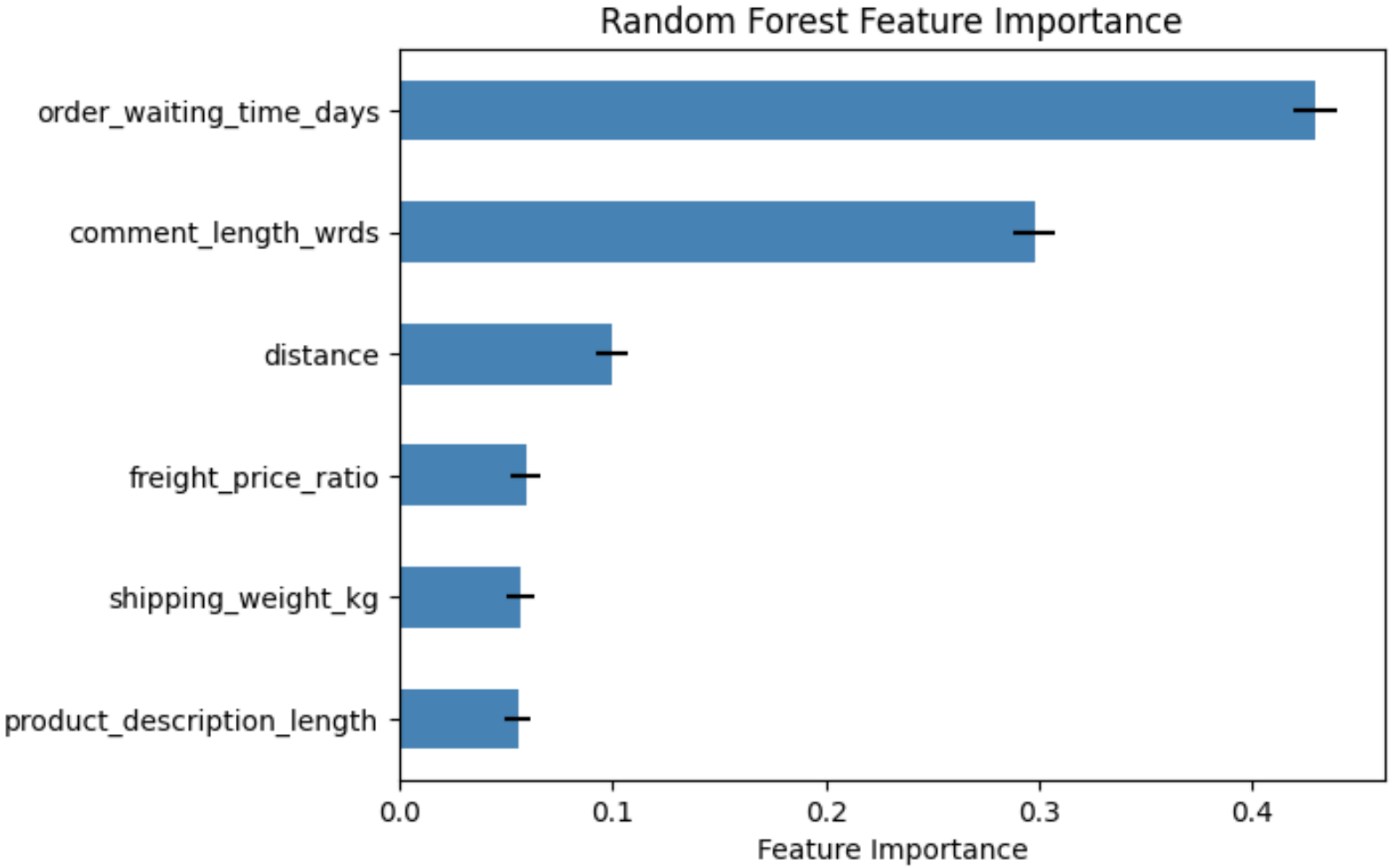
Let's try and explain waiting times and review score impact better



Distance has a ***negative*** impact

Negative impact of weight and size

Supervised machine learning

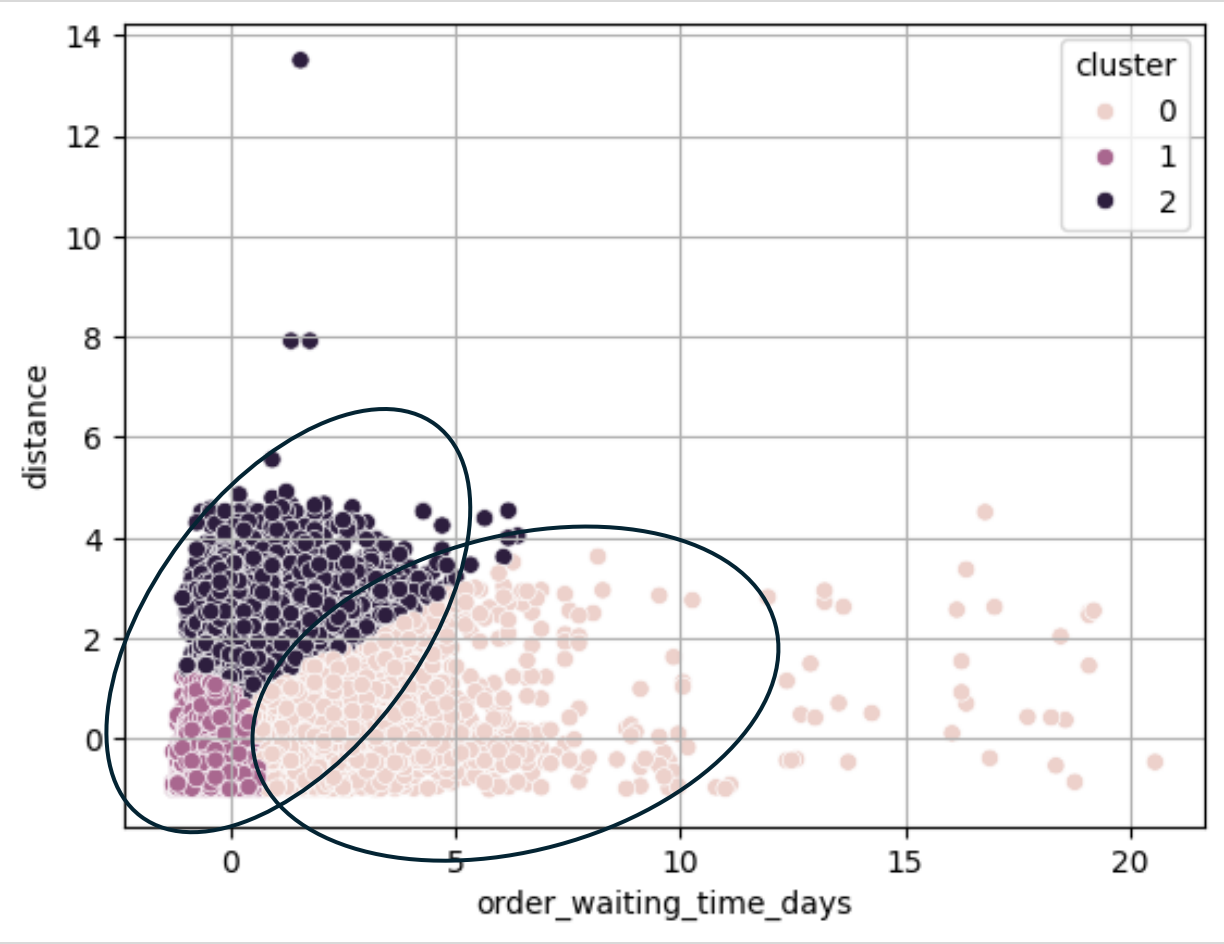


Shipping distance and time ***strongly linked*** to customer satisfaction!

How does it affect customer satisfaction?

Review impact clusters (scaled factors):

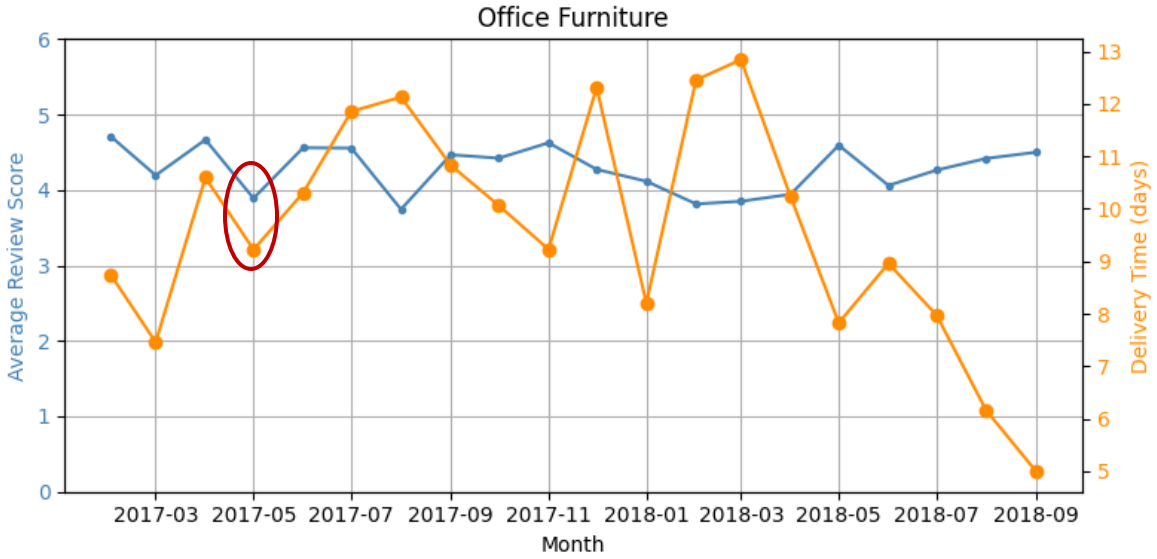
Unsupervised machine learning



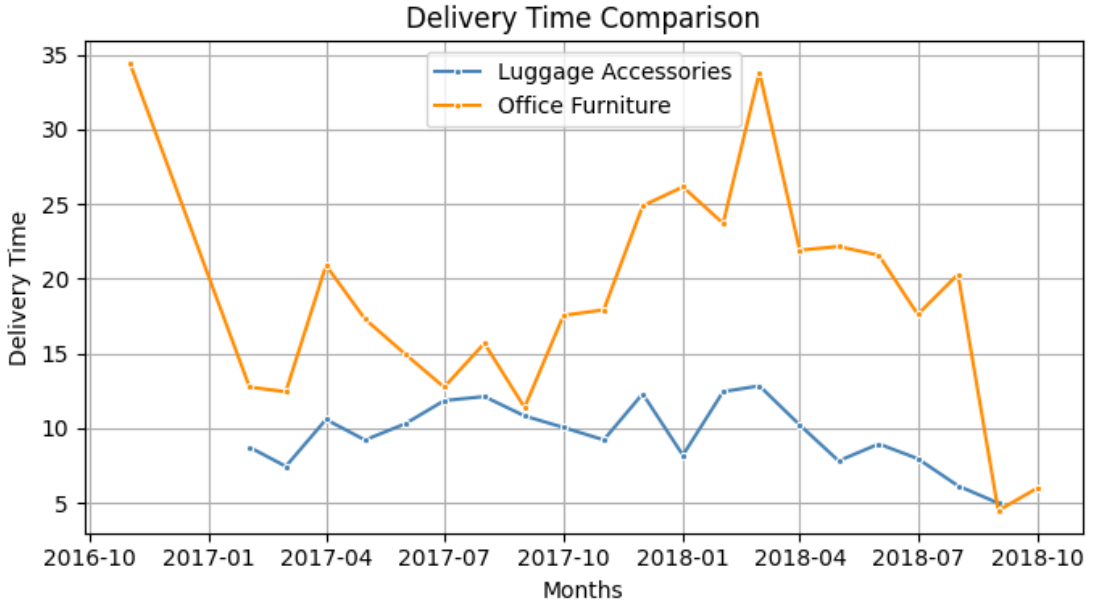
Delivery | Distance	Quick	Average	Slow
Close	Very satisfied (4.4)		
In between			Not satisfied (3.3)
Far		Satisfied (4.1)	

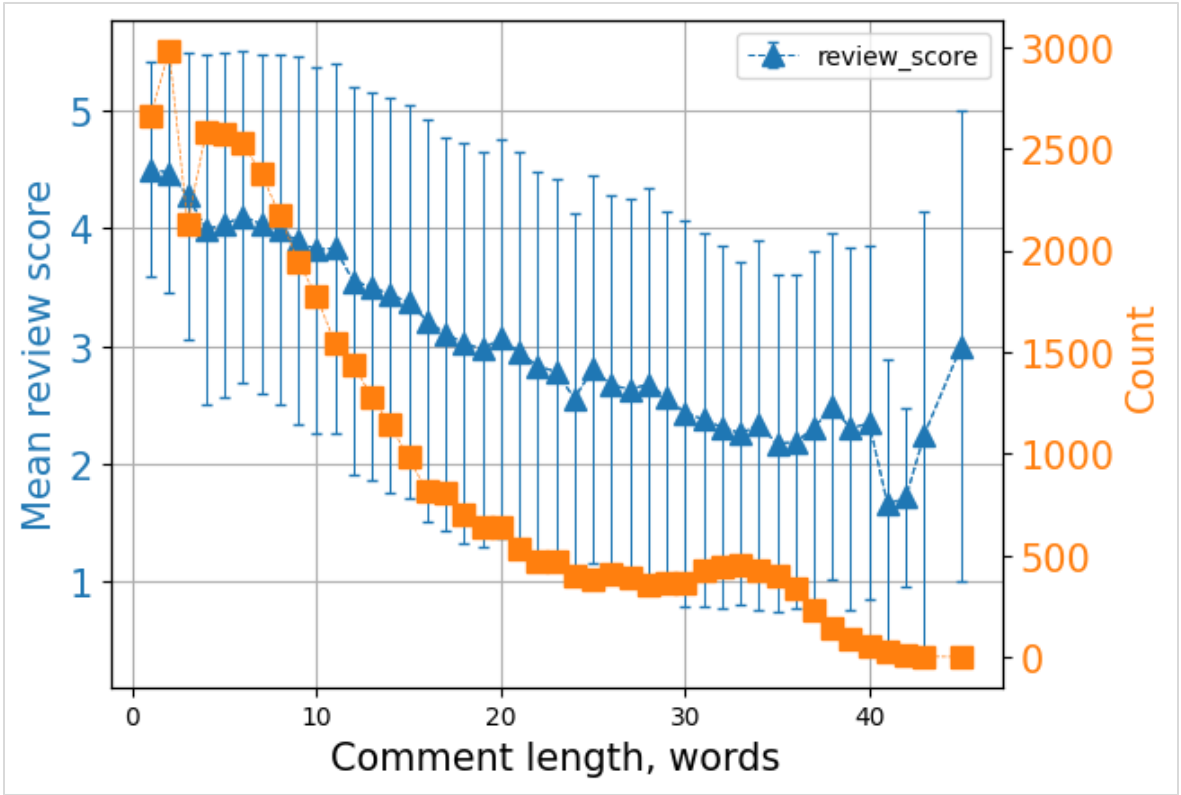
What drives customer satisfaction?

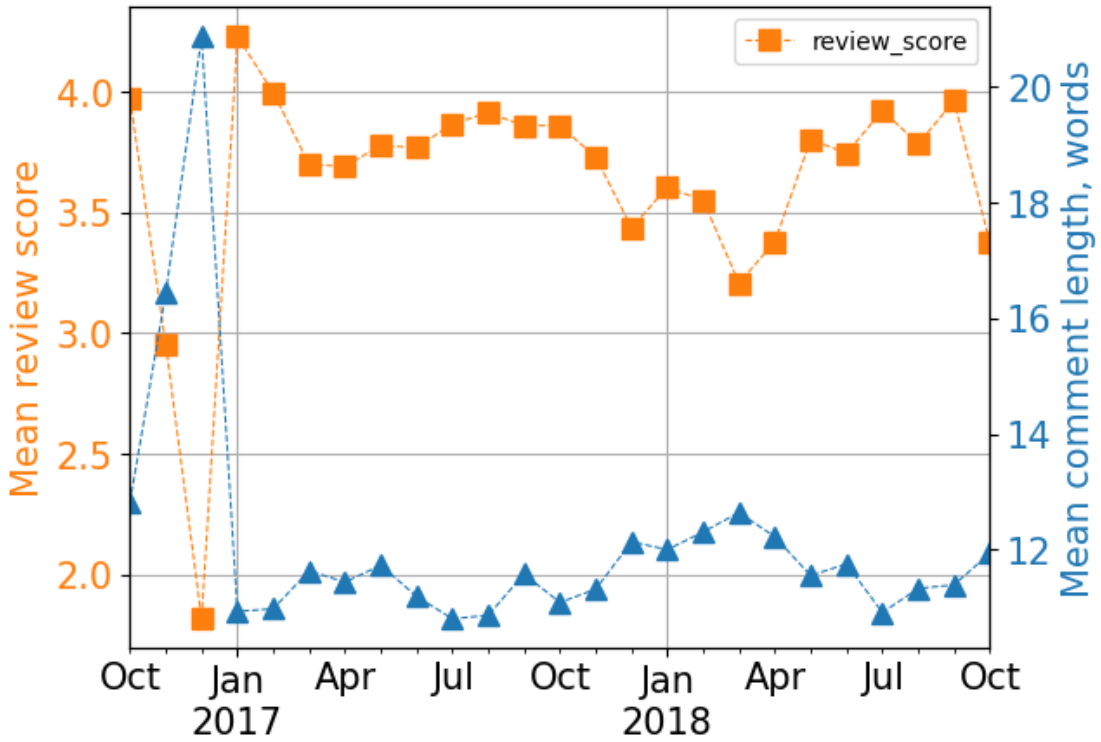
Analysis suggests shipping time and distance



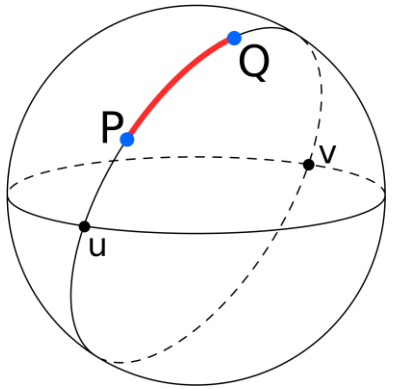
Maybe a quality issue




review_score
Mean review score
Count
Comment length, words
0
10
20
30
40
1
2
3
4
5
0
500
1000
1500
2000
2500
3000


review_score
Mean review score
Mean comment length, words
Oct
Jan
2017
Apr
Jul
Oct
Jan
2018
Apr
Jul
Oct
2.0
2.5
3.0
3.5
4.0
12
14
16
18
20

Seller - customer distance



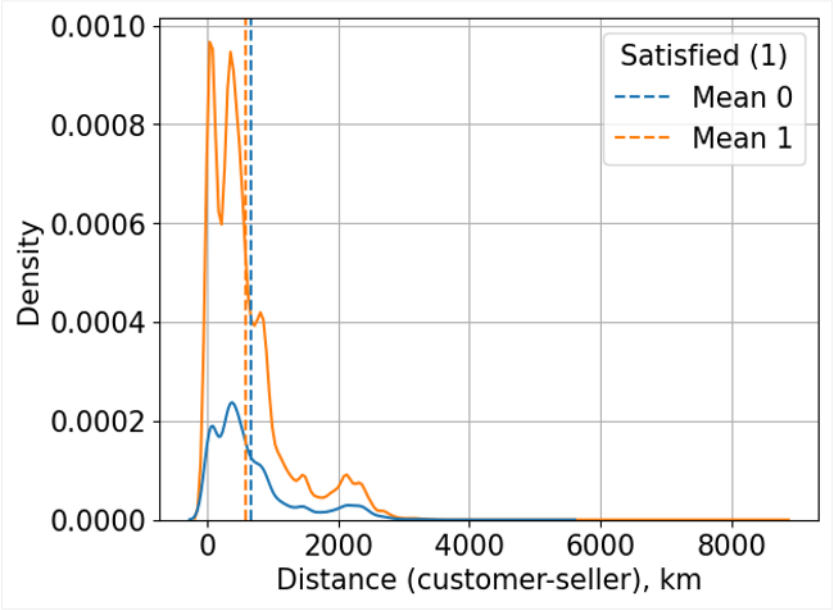
Geolocation dataset

- ~10E6 rows, 5 columns (e.g., zip code, longitude, latitude, city, state)
  - ~19K zip codes (first 5 digits only)
  - ~700K lat./long. Coordinates
- Grouped by zipcode and averaged lat./long.
- Merged left first on seller zipcodes, merged left second on customer zipcodes

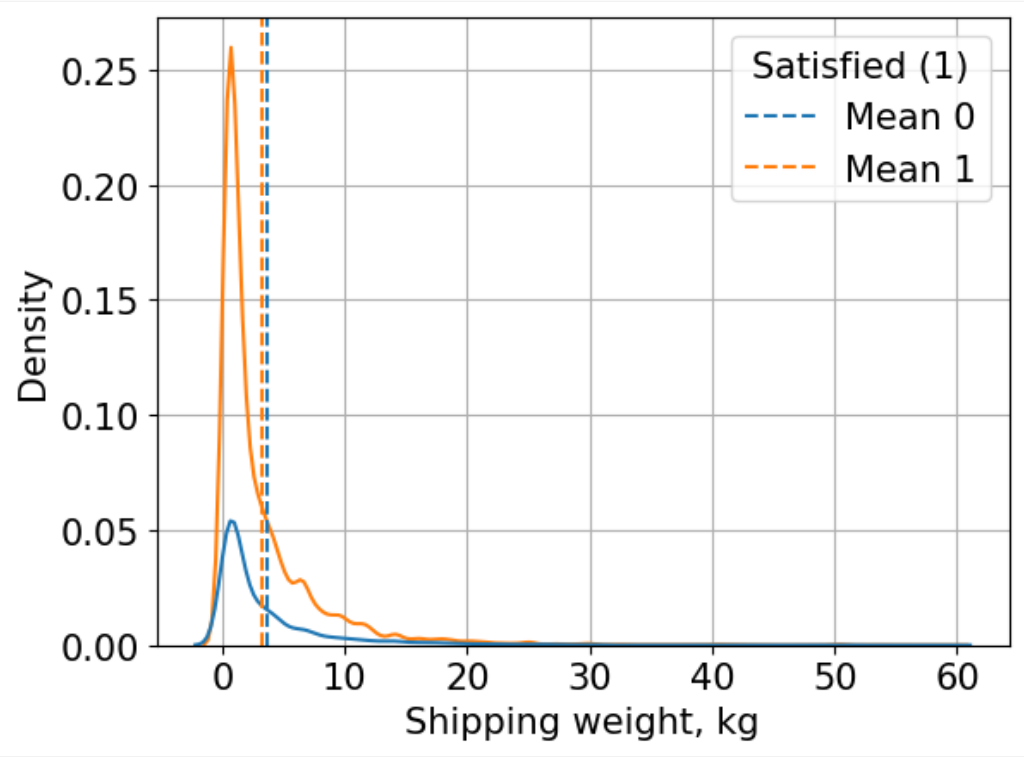
By examining the review scores as 0 (not satisfied) - 1(satisfied)

Haversine formula

- Shortest distance [km] between two points (e.g., P, Q) on a sphere knowing longitudes and latitudes
  - P ~ seller, Q ~ customer



Let's try to explain better the waiting time and its impact



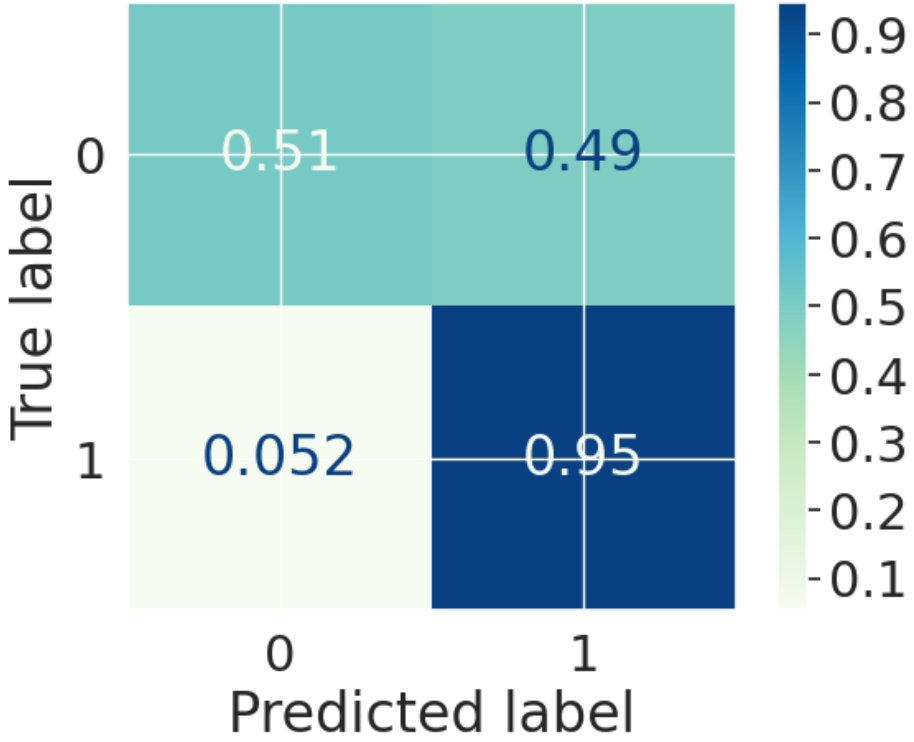
By examining the review scores as 0 (not satisfied) - 1(satisfied)

Shipping weight - kg:

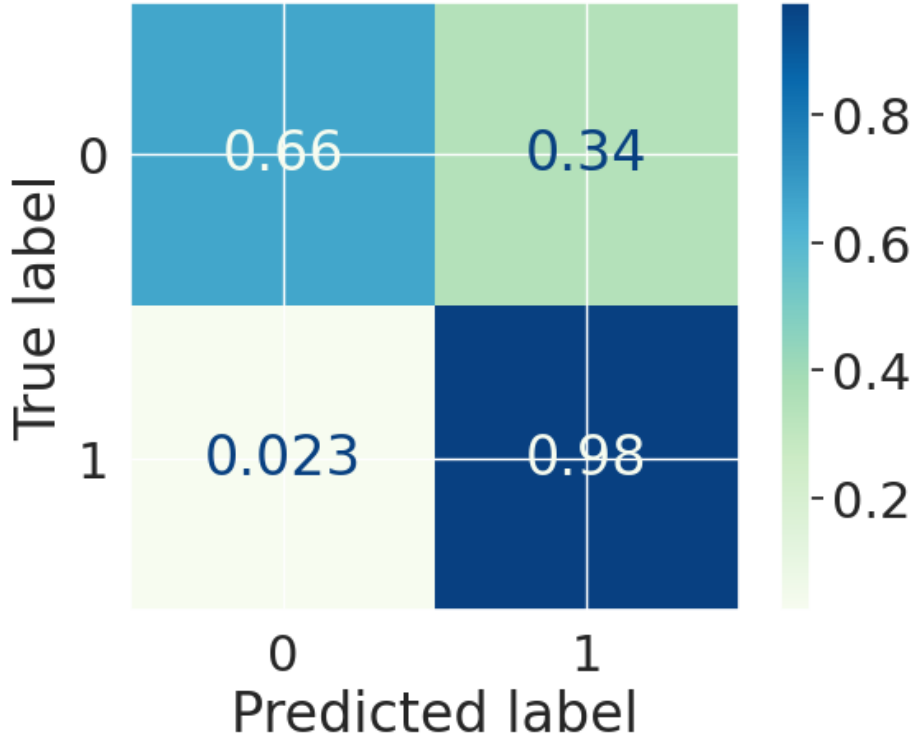
- Volumetric weight, kg=(width x height x length)/5000 for ground
- Carriers use higher of volumetric weight or physical weight – shipping weight - kg

Supervised machine learning: Classifiers

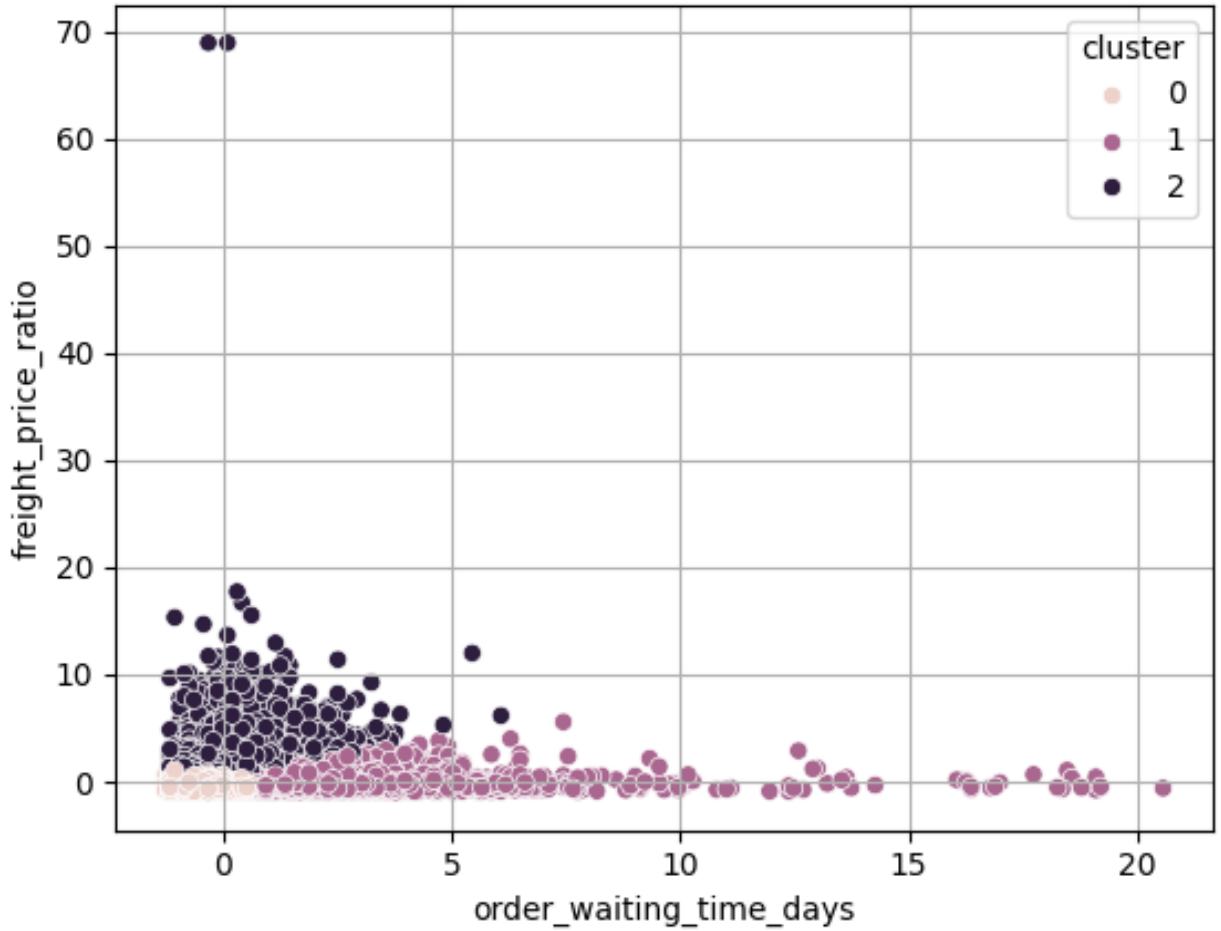
Random Forest



Catboost



Total delivery time + freight/price

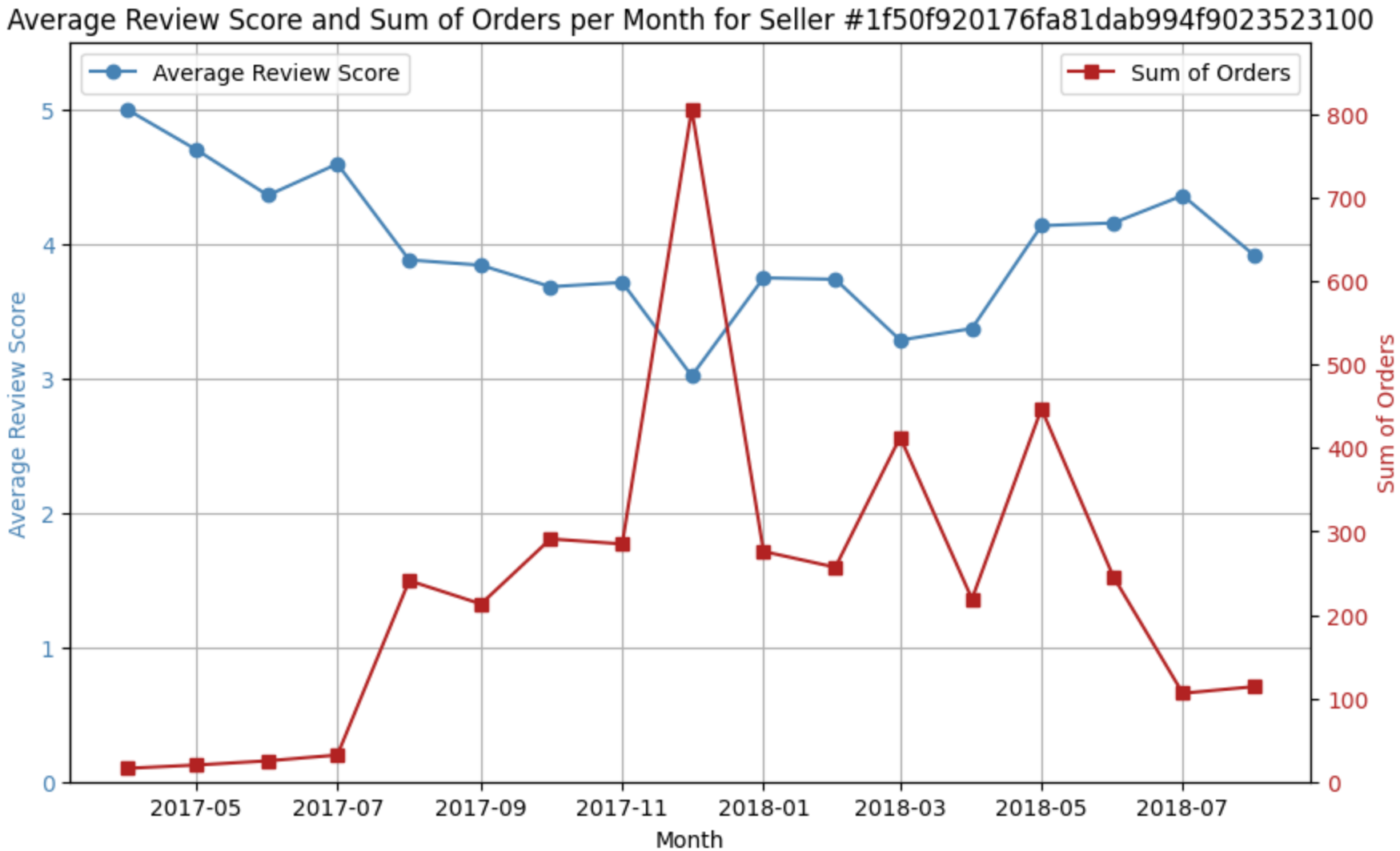


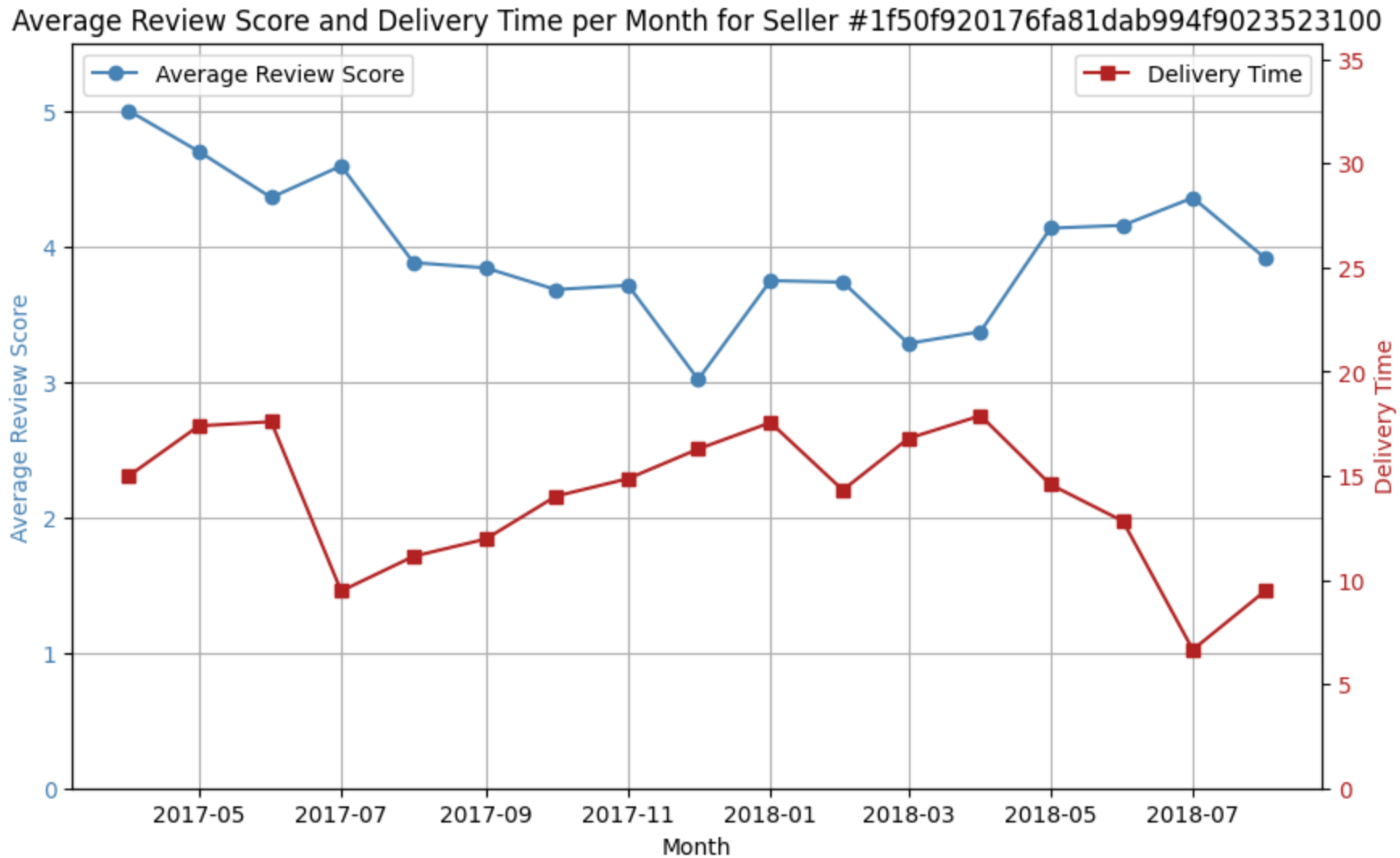
cluster	mean review
0	4.390260
2	4.266413
1	3.329732

Does Lannister always pay his debts?

The OLIST story

- 2.85


Average Review Score and Sum of Orders per Month for Seller #1f50f920176fa81dab994f9023523100
Average Review Score
Sum of Orders
Average Review Score
Sum of Orders
Month
0
1
2
3
4
5
0
100
200
300
400
500
600
700
800
2017-05
2017-07
2017-09
2017-11
2018-01
2018-03
2018-05
2018-07


Average Review Score and Delivery Time per Month for Seller #1f50f920176fa81dab994f9023523100
Average Review Score
Delivery Time
Average Review Score
Delivery Time
Month
0
1
2
3
4
5
0
5
10
15
20
25
30
35
2017-05
2017-07
2017-09
2017-11
2018-01
2018-03
2018-05
2018-07