

How the consumer behavior could be described for the Olist Platform (Brazilian eCommerce)?

Xlmena Rios Cotazo, Ramon Manuel Sandoval, Luisa Maria Carabali, Alejandro Camargo, Cristian Sarmiento, Hector Melo



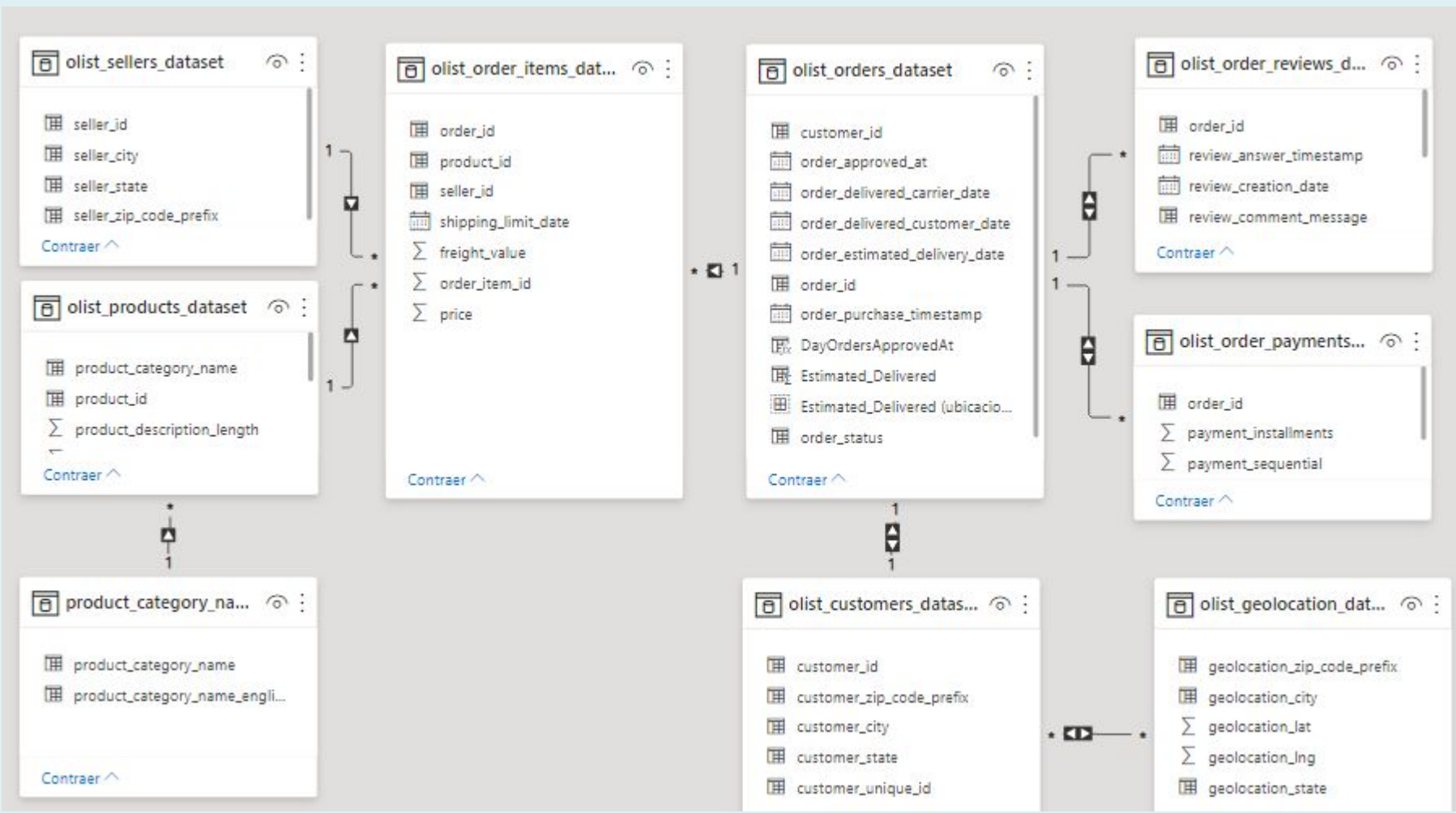
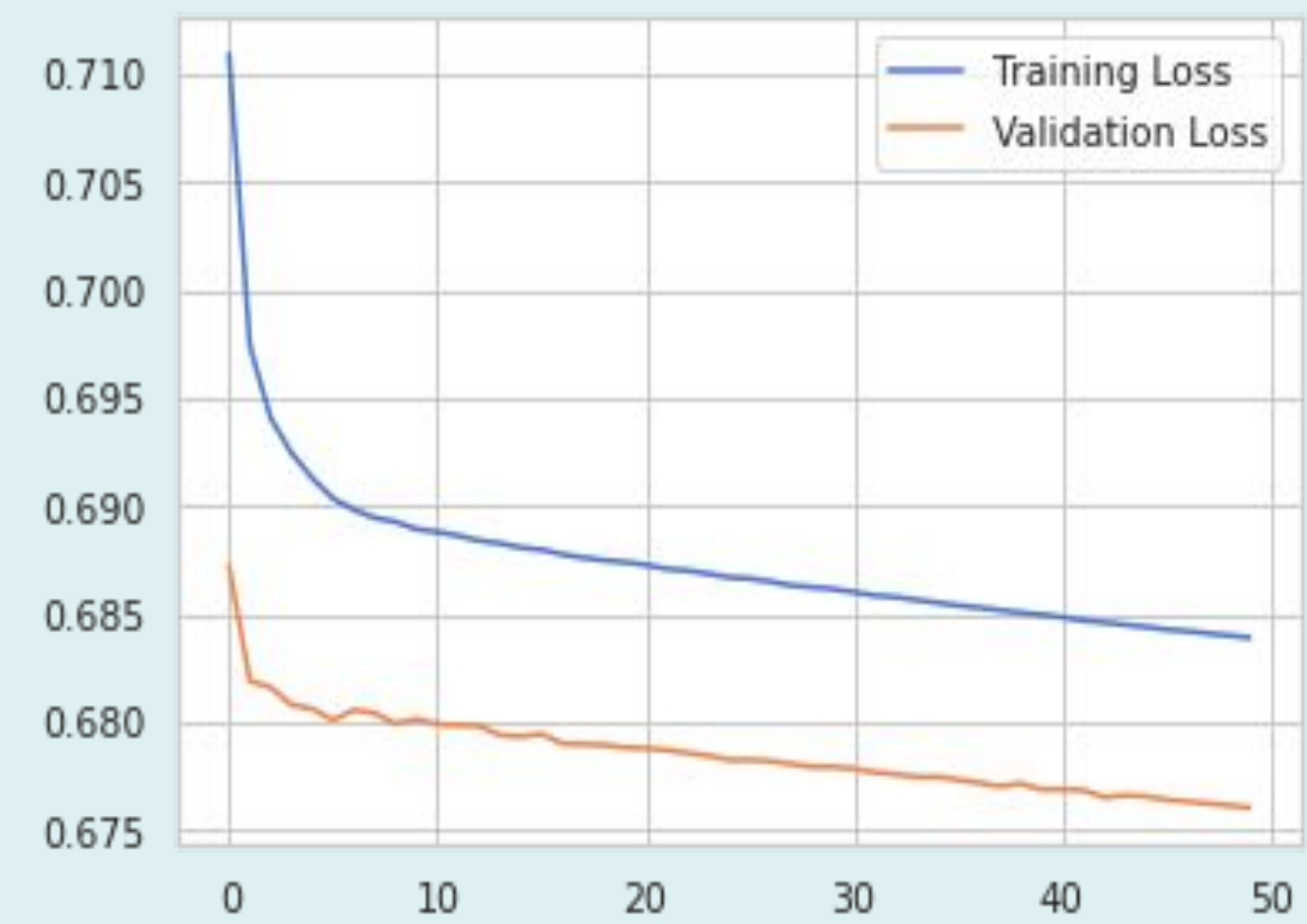
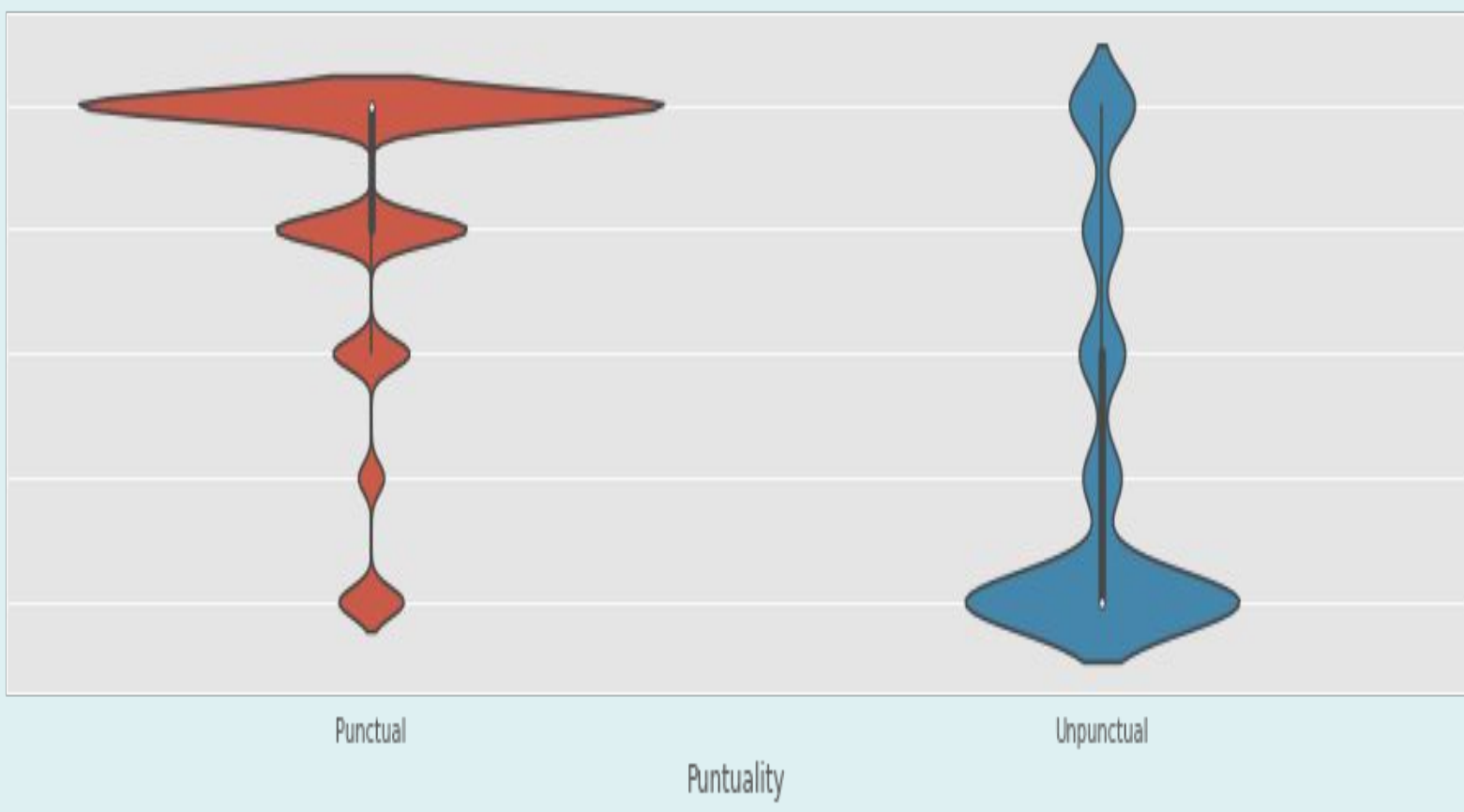
Background

What aspects of consumers online behavior is useful for businesses to better understand their customers and predict consumer demand

Almost **60 %** of orders with late deliveries got a 1 star review

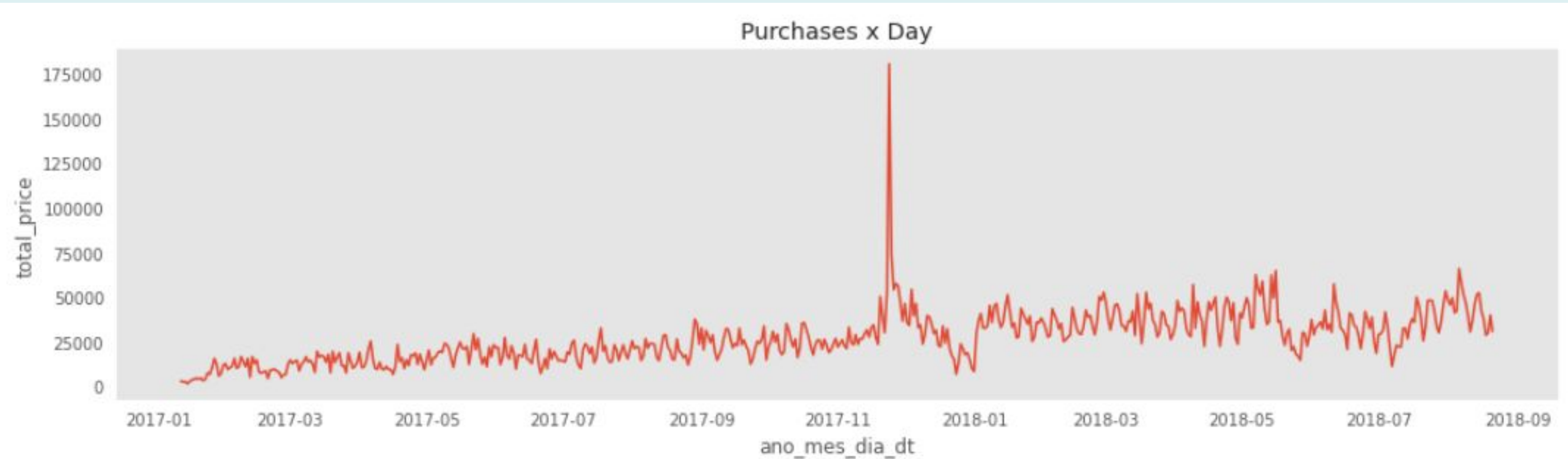
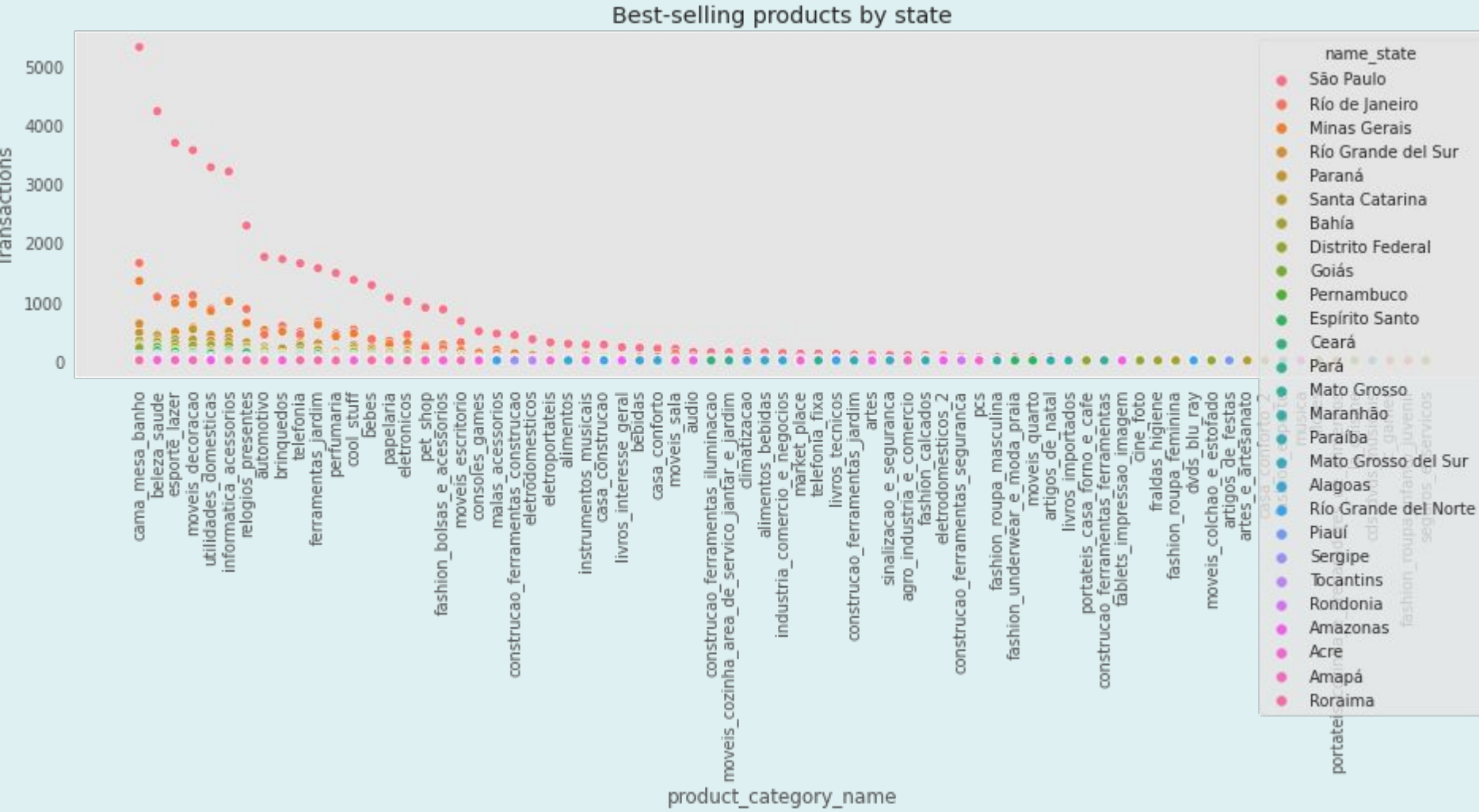
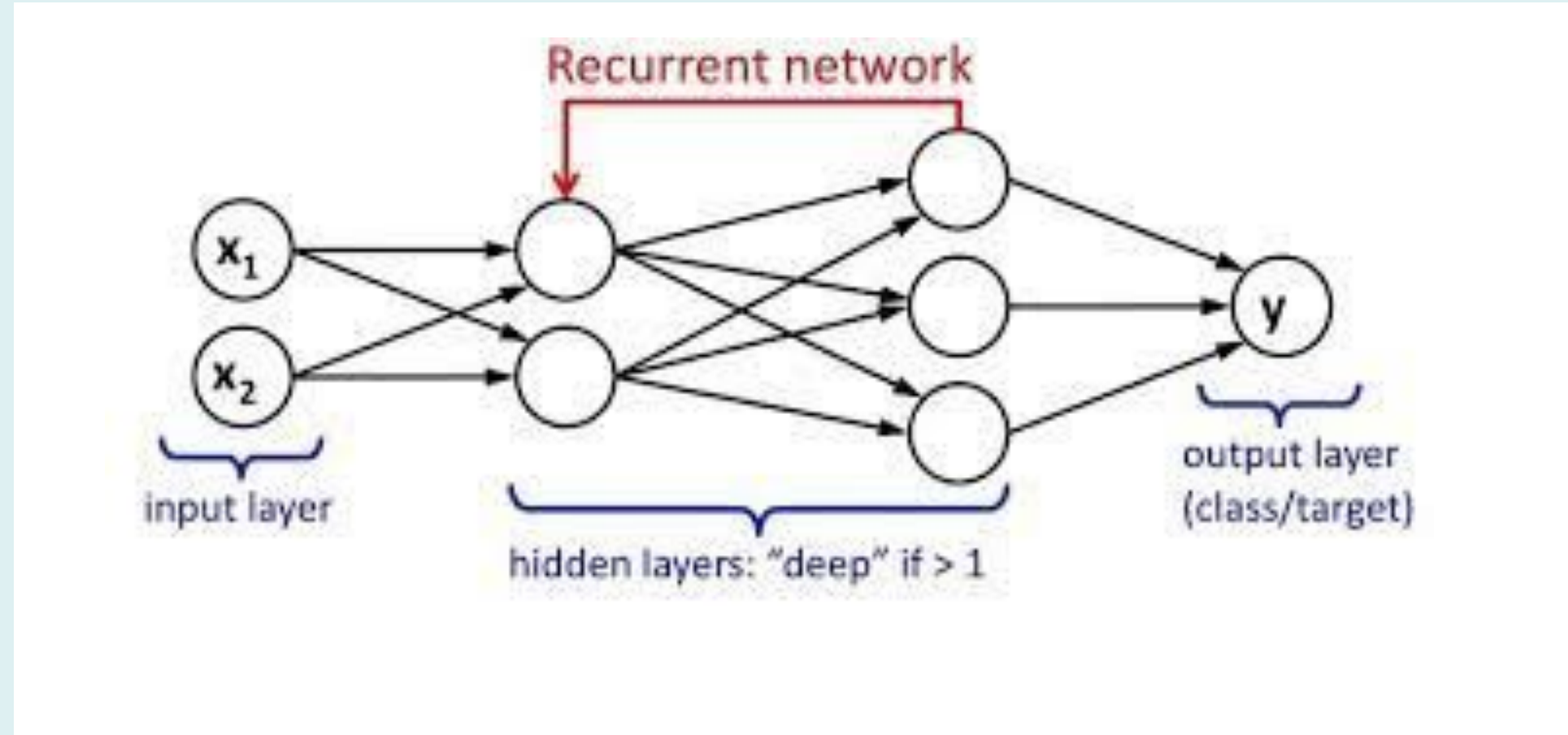
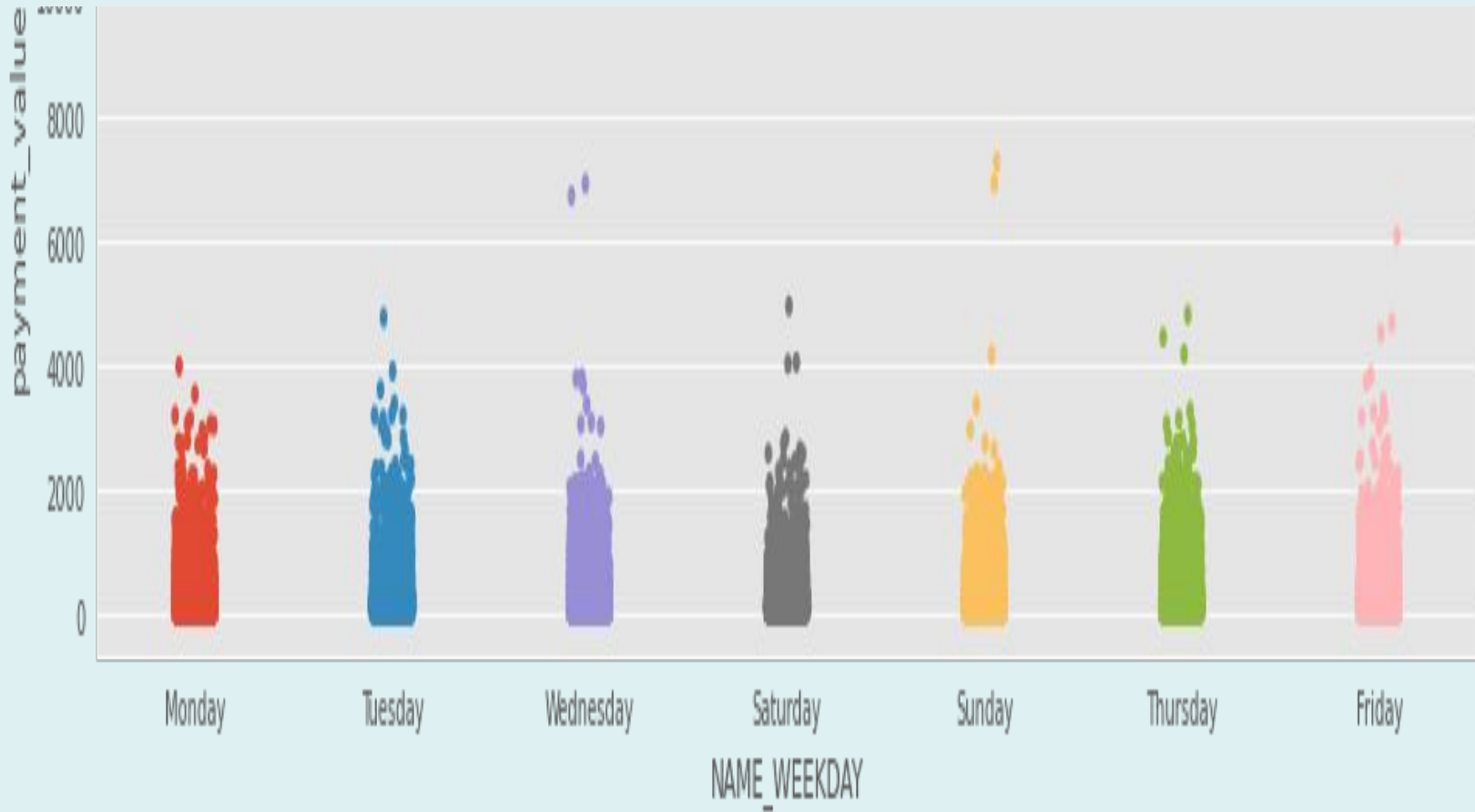
5.653/9.550

Exploratory Data Analysis (EDA)



Data

we analyze a Brazilian ecommerce public dataset of orders made at **Olist Store**. The dataset has information of **100k orders from 2016 to 2018** made at multiple marketplaces in Brazil.



EDA Analysis

- From the EDA we found out the states that make the most purchases are Sao Paulo and Rio de Janeiro, the two main states in the country.
- Almost **60 %** of orders with **late deliveries** got a score of **1 star** in the **product review**.
- December is the most important month in total sales.
- Most people in Brazil use **Credit Card** to pay their online purchases.
- The categories of products with more sales are: Bed table bath, health and beauty, sports, and decoration furniture.
- Purchasing by hour is low from 12:00AM, and starts decreasing till 5:00 AM.

Analysis/Modelling



Models/Insights



Sales Count by Month for Category

OLS Regression Results					
Dep. Variable:	total_sales	R-squared:	0.970		
Model:	OLS	Adj. R-squared:	0.968		
Method:	Least Squares	F-statistic:	644.8		
Date:	Sat, 19 Dec 2020	Prob (F-statistic):	1.09e-16		
Time:	19:51:50	Log-Likelihood:	-115.60		
No. Observations:	22	AIC:	235.2		
Df Residuals:	20	BIC:	237.4		
Df Model:	1				
Covariance Type:	nonrobust				
	coef	std err	t	P> t	[0.025 0.975]
Intercept	-19.9722	20.941	-0.954	0.352	-63.655 23.710
month_codes	40.5975	1.595	25.392	0.000	37.180 43.835
Omnibus:	0.018	Durbin-Watson:	1.860		
Prob(Omnibus):	0.991	Jarque-Bera (JB):	0.100		
Skew:	-0.013	Prob(SB):	0.951		
Kurtosis:	2.670	Cond. No.	26.7		

What does affect the total price?

OLS Regression Results					
Dep. Variable:	total_price	R-squared:	0.156		
Model:	OLS	Adj. R-squared:	0.156		
Method:	Least Squares	F-statistic:	2287.		
Date:	Fri, 18 Dec 2020	Prob (F-statistic):	0.00		
Time:	21:01:07	Log-Likelihood:	-7.3106e+05		
No. Observations:	111170	AIC:	1.462e+06		
Df Residuals:	111168	BIC:	1.462e+06		
Df Model:	9				
Covariance Type:	nonrobust				
	coef	std err	t	P> t	[0.025 0.975]
Intercept	-1.225e+04	2087.631	-5.868	0.000	-1.63e+04 -8158.519
regions[T.Sur]	-52.0669	4.091	-12.727	0.000	-60.086 -44.048
regions[T.Sudeste]	-61.5766	3.984	-15.771	0.000	-69.229 -53.924
regions[T.Nordeste]	-13.5198	4.215	-3.207	0.001	-21.782 -5.258
regions[T.CentroOeste]	-37.3676	4.410	-8.473	0.000	-46.012 -28.723
year	6.1383	1.035	5.932	0.000	4.110 8.166
product_weight_g	0.0151	0.000	64.904	0.000	0.015 0.016
product_vol	0.0008	3.73e-05	21.378	0.000	0.001 0.001
product_photos_qty	4.5273	0.303	14.934	0.000	3.933 5.121
review_score	1.3929	0.376	3.705	0.000	0.650 2.130

Conclusions

- Promoting quality and quantity of product photos with sellers could be a good strategy to improve sales, according to models any changes in this variable have a positive impact on the total sales.
- Review score is key in order to increase sales, promoting among customers the evaluation of their orders is a good practice.
- Recurrent Neural Networks are designed to time series modeling, for our case we build a model to predict sales according to the historical data.