



The Business Game

NTTdata x LUISS

Cosimo Poccianti
Enrico Grandi
Miro Confalone
Chi Quy Loc

Final Presentation

TO BETTER VISUALISE THE
INTERACTIVE PART OF THE
PRESENTATION, WE RECOMMEND
VIEWING IT AT THE FOLLOWING
LINK



Click here

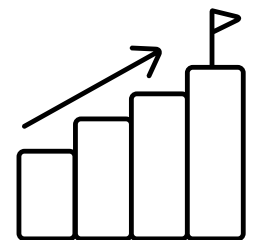


[https://www.canva.com/design/
DAFAAn52twU/X2xUK3zTAV0y6
RwziYRYyA/view?
utm_content=DAFAAn52twU&
utm_campaign=designshare&ut
m_medium=link&utm_source=p
ublishsharelink](https://www.canva.com/design/DAFAAn52twU/X2xUK3zTAV0y6RwziYRYyA/view?utm_content=DAFAAn52twU&utm_campaign=designshare&utm_medium=link&utm_source=publishsharelink)

INDEX

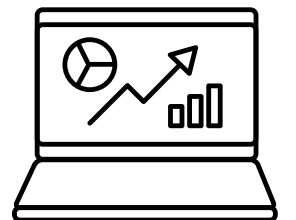
- Business KPIs
- Customer Segmentation
- Product recommendation

- Business actions
- Shipping efficiency



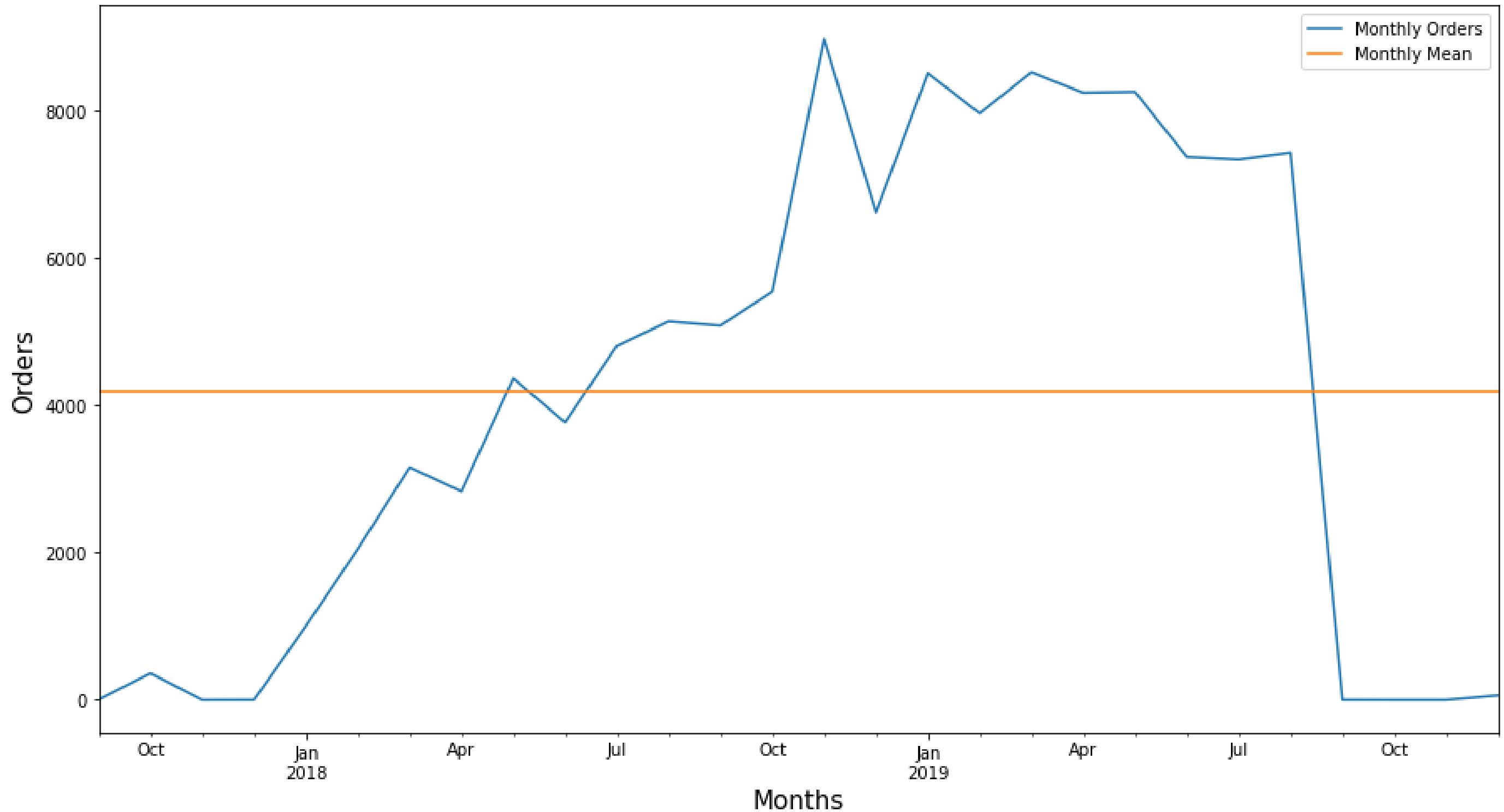
Business KPIs

1. Daily & Weekly Orders
2. Items per Order
3. Efficiency (Delivery Delay)
4. Best Sellers & Buyers
5. Best Lines of Products



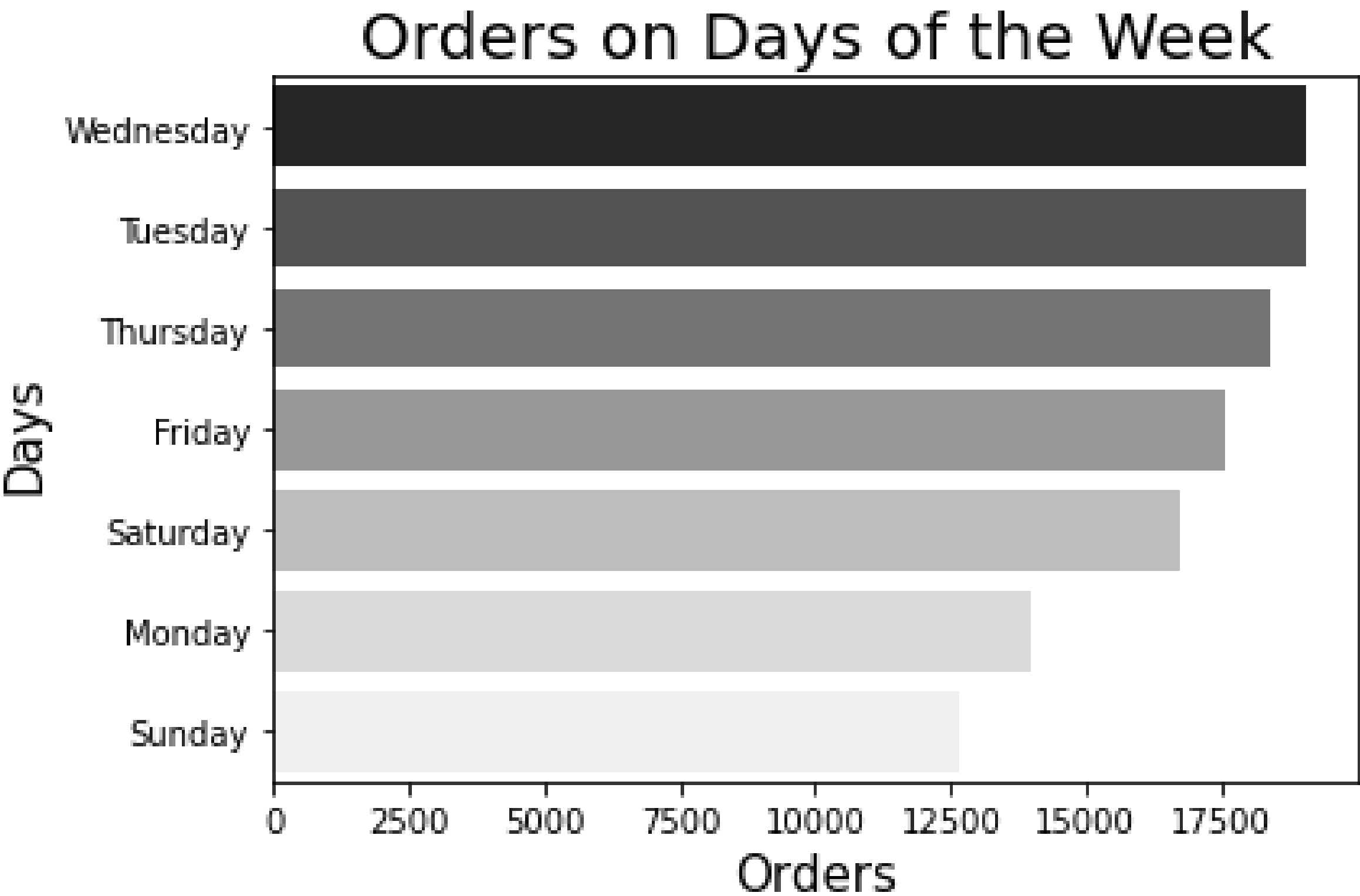
Monthly Orders

orders	
count	28
mean	4,193
std	3,348
min	0
25%	285
50%	4,586
75%	7,386
max	8,977



Orders per weekdays

day_of_week	mean
Sunday	146
Monday	159
Saturday	190
Friday	197
Thursday	212
Tuesday	214
Wednesday	219



Most Valuable Customers

			amount_paid	orders_purchased
			sum	sum
customer_autonomous_community	customer_city	customer_unique_id		
Cantabria	Santander	0f75637a7e2f7b6ae43665d1f7d0ca5d	304,412	1
País Vasco	San Sebastián	ff0ae98646e7bbb41cf0f0d3991fef98	304,827	1
Islas Canarias	Las Palmas de Gran Canaria	58c1b085b54c03a1f1ab5f13d64c2b1c	306,472	2
Cataluña	Gerona	c6111f70f40b3420e387493c627c27fa	312,650	1
Galicia	Coruña	895617ab63a9ad8881d9470f7427cd25	312,650	1
Castilla y León	Valladolid	46450c74a0d8c5ca9395da1daac6c120	318,434	1
Andalucía	Sevilla	2c3b08cf3584d8c0a81d9618a73aa219	352,646	1
Región de Murcia	Murcia	5d09b0d82126457e2a8ebfb9c9a1ffc4	373,622	1
Comunidad de Madrid	Alcobendas	fa562ef24d41361e476e748681810e1e	417,526	1
	Madrid	edde2314c6c30e864a128ac95d6b2112	451,332	1

Best Products

		gross_revenue	review_score	shipping_cost
		sum	mean	mean
product_category_name	product_id			
computer accessories	470433f95ba906e17efac3fce39e9ffd	1,520,056	3	4,022
videogame console	fb01a5fc09b9b9563c2ee41a22f07d54	1,611,537	4	3,054
beauty & personal care	7c1bd920dbdf22470b68bde975dd3ccf	1,759,179	4	1,515
comics	601a360bd2a916ecef0e88de72a6531a	1,999,633	4	1,830
automotive	fd0065af7f09af4b82a0ca8f3eed1852	2,034,134	4	3,414
comics	165f86fe8b799a708a20ee4ba125c289	2,212,669	4	2,201
beauty & personal care	19c91ef95d509ea33eda93495c4d3481	2,422,266	4	3,419
	2b4609f8948be18874494203496bc318	2,683,981	4	1,462
computers tablets	d6160fb7873f184099d9bc95e30376af	2,945,667	4	3,956
computer accessories	d1c427060a0f73f6b889a5c7c61f2ac4	3,601,205	4	3,646

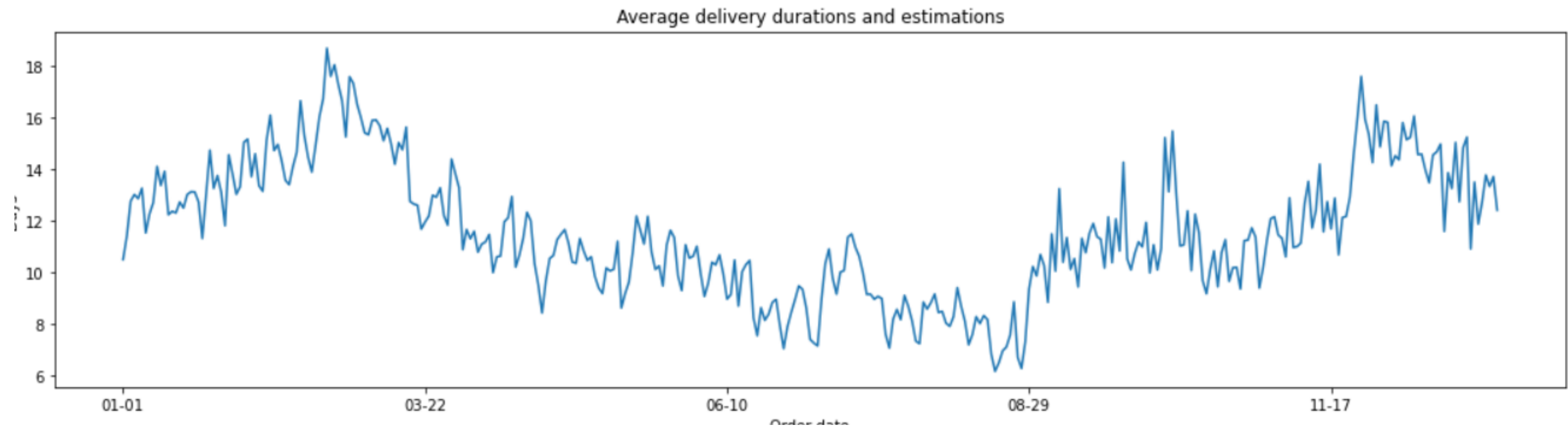
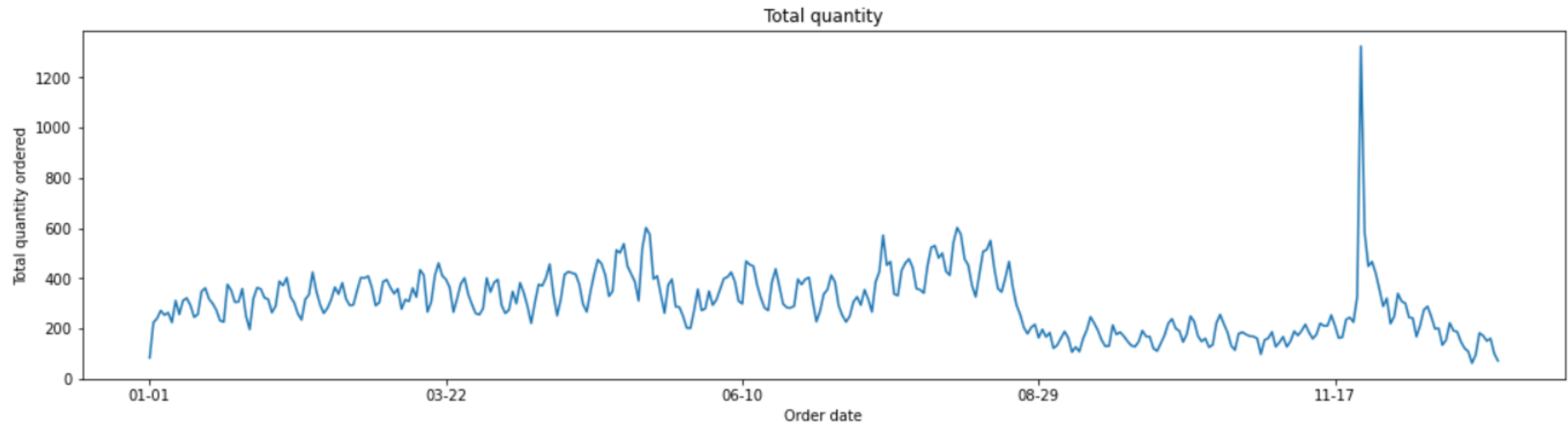
Best Sellers per revenues

seller_autonomous_community	seller_city	seller_id	gross_revenue	review_score
			sum	mean
Andalucía	Cádiz	c826c40d7b19f62a09e2d7c5e7295ee2	4,750,886	4
País Vasco	San Sebastián	53243585a1d6dc2643021fd1853d8905	6,099,050	4
Comunidad de Madrid	Madrid	7ddcbb64b5bc1ef36ca8c151f6ec77df	6,181,046	4
	Parla	a1043baf471dff536d0c462352beb48	7,088,849	4
Cataluña	Tarragona	edb1ef5e36e0c8cd84eb3c9b003e486d	7,893,316	4
Andalucía	Córdoba	7e93a43ef30c4f03f38b393420bc753a	11,631,884	4
		cc419e0650a3c5ba77189a1882b7556a	12,166,208	4
Cataluña	Barcelona	46dc3b2cc0980fb8ec44634e21d2718e	12,955,776	4
Comunidad de Madrid	Madrid	7a67c85e85bb2ce8582c35f2203ad736	15,822,780	4
Andalucía	Córdoba	7c67e1448b00f6e969d365cea6b010ab	16,151,865	4

Most Delays Sellers

			delay	review_score
			count	mean
seller_autonomous_community	seller_city	seller_id		
Andalucía	Almería	da8622b14eb17ae2831f4ac5b9dab84a	95	4
Comunidad de Madrid	Madrid	8b321bb669392f5163d04c59e235e066	96	4
Andalucía	Córdoba	7c67e1448b00f6e969d365cea6b010ab	96	4
Comunidad de Madrid	Madrid	955fee9216a65b617aa5c0531780ce60	96	4
Andalucía	Córdoba	cc419e0650a3c5ba77189a1882b7556a	98	4
Comunidad de Madrid	Madrid	6560211a19b47992c3666cc44a7e94c0	115	4
		ea8482cd71df3c1969d7b9473ff13abc	117	4
	Sevilla	4869f7a5dfa277a7dca6462dcf3b52b2	128	4
Andalucía	Huelva	1f50f920176fa81dab994f9023523100	145	4
	Granada	4a3ca9315b744ce9f8e9374361493884	188	4

Visualization of total orders per day and average number of days per delivery



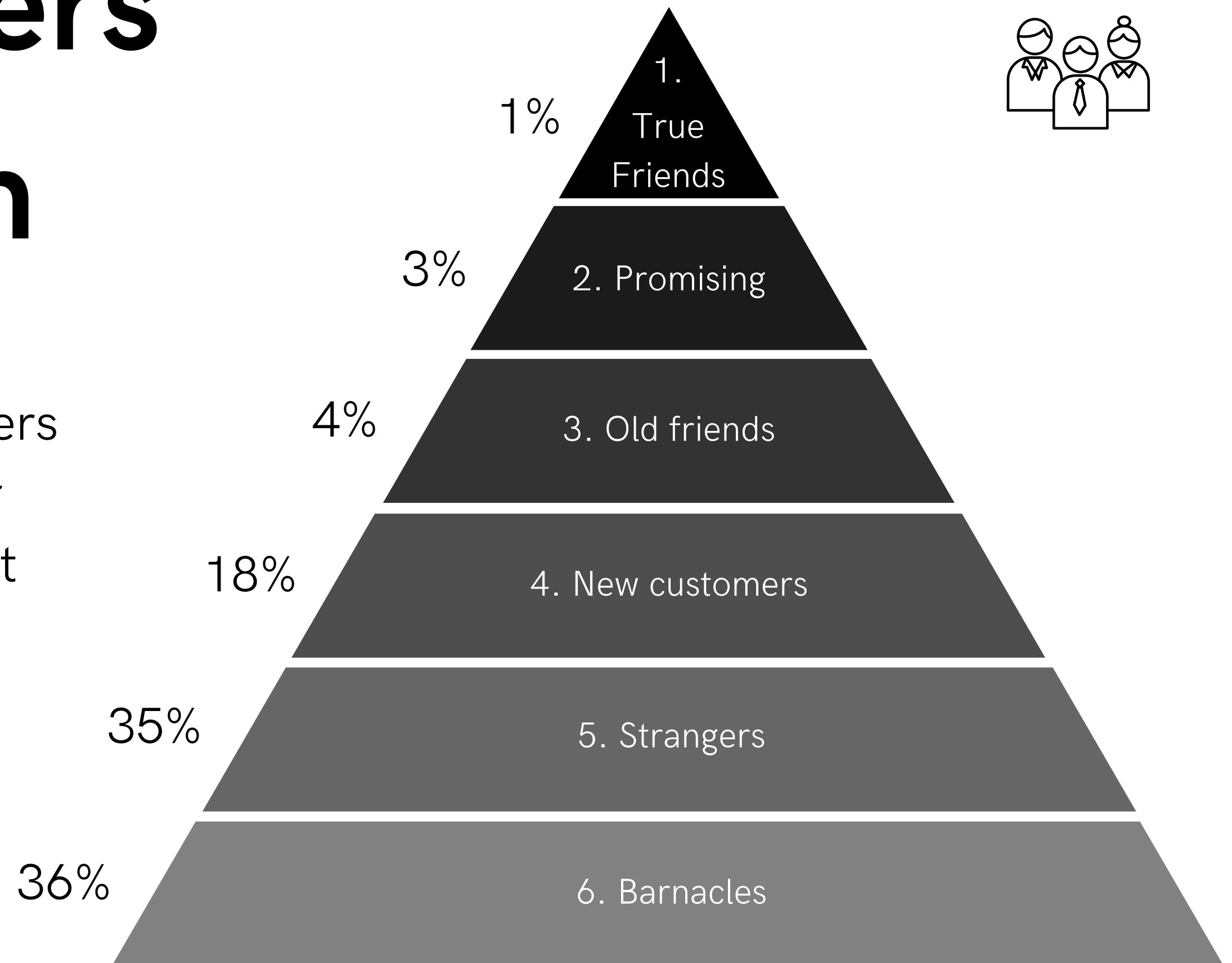
Number of customer per City

It is possible to see which city the most customers come from



RFM customers segmentation

Customers were divided into clusters according to the frequency of their purchases and the date of their last purchase.



CLUSTERING SEGMENTATION

DIVISION OF CUSTOMERS INTO CLUSTERS WITH K-MEANS

Four different clusters were identified, and classified according to the number of orders placed, the last date of purchase and the amount of expenditure

Strangers

People that made few orders and long time ago with low spending, Cluster 1

Barnacles

Good Recency, low frequency, and second lowest spending, Cluster 0

Loyals

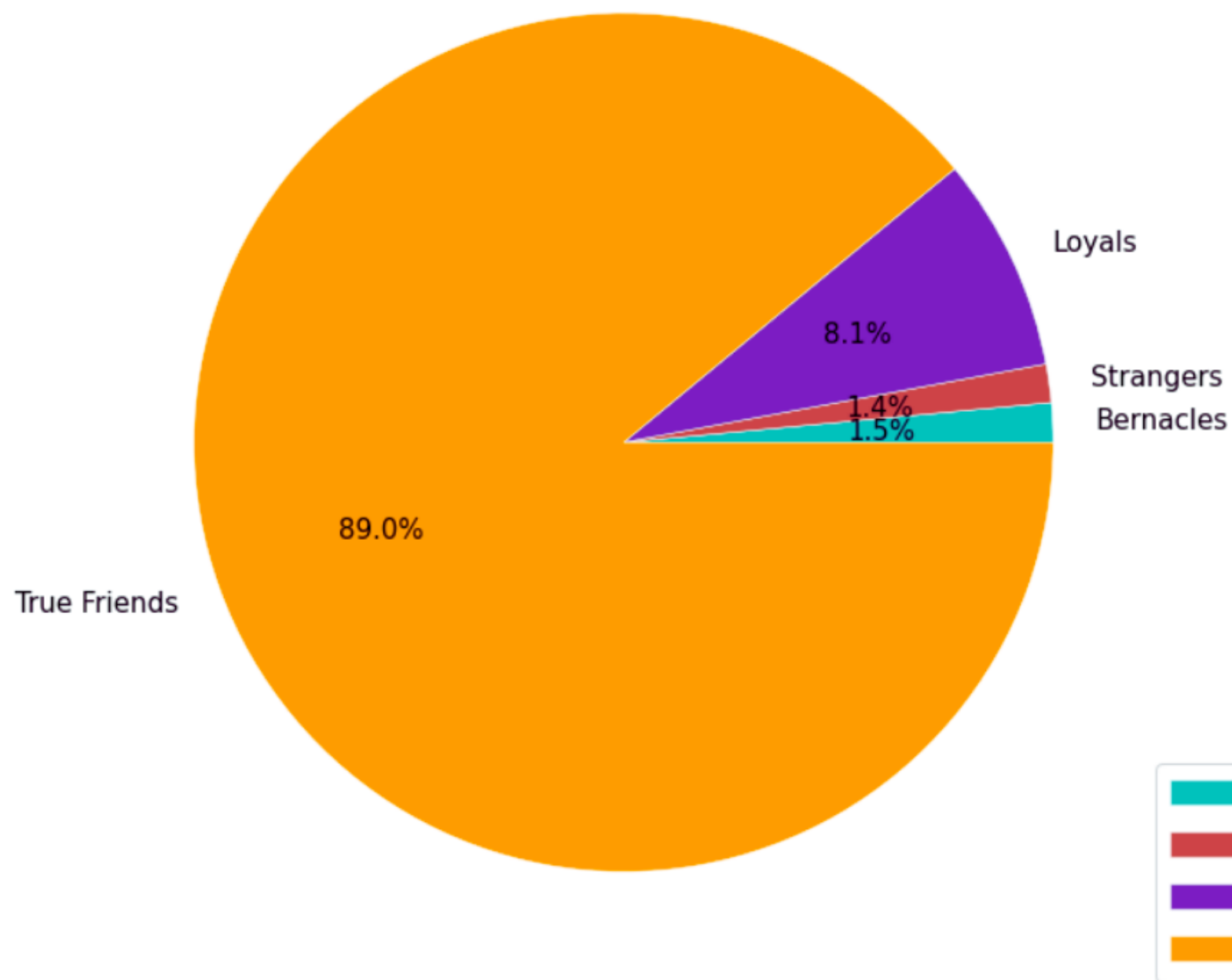
customer that have important amounts, are pretty recent, and medium to high budget, Cluster 2

True Friends

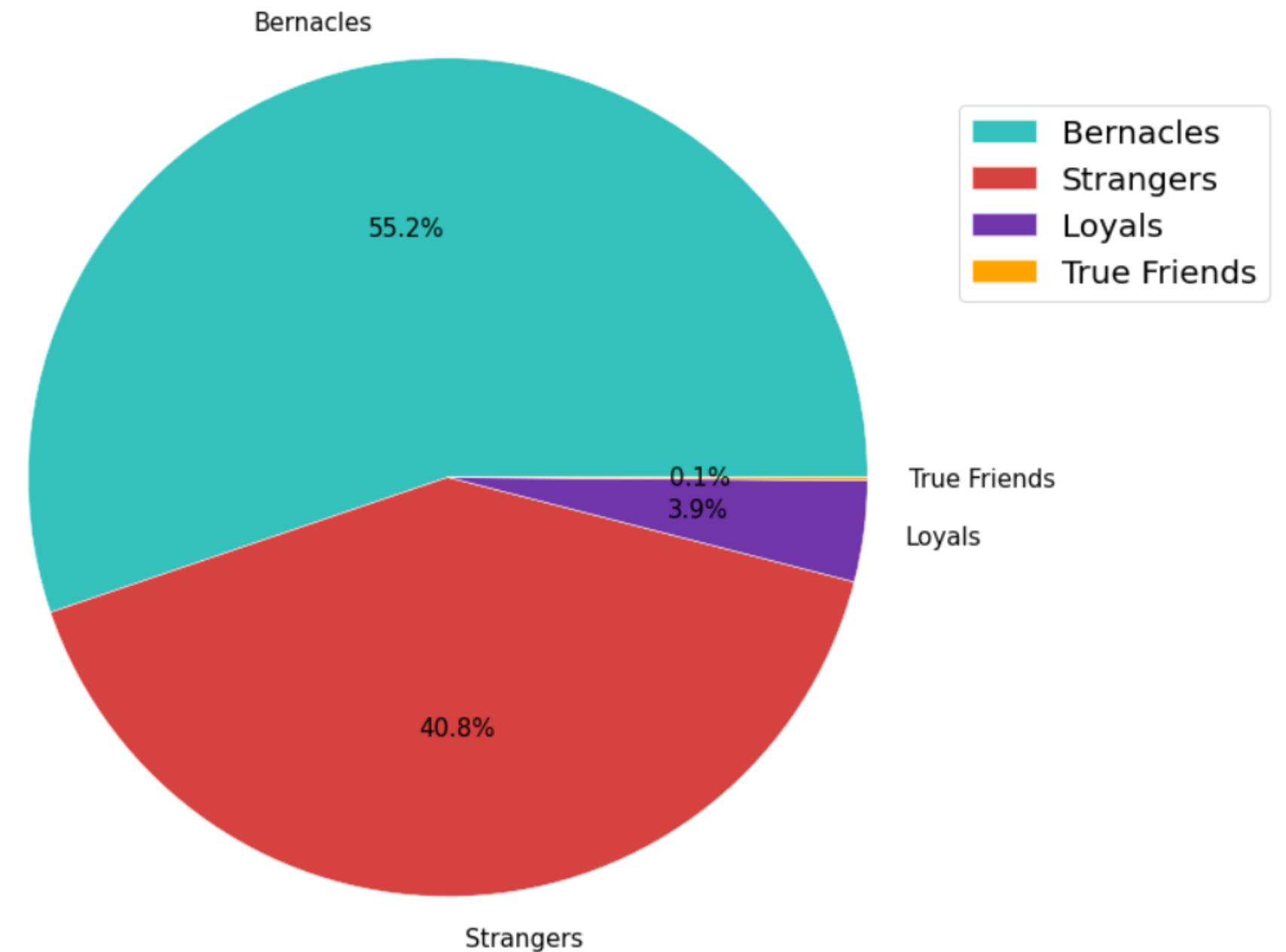
People that spends often, frequently and for the highest amounts, Cluster 3

0,1% of customer are segmented as True friend, but they generate the 89% of the revenue

Average revenue per K-Means cluster



Customers per K-Means cluster



Average cluster value per city

It is possible to visualize the more common customers segment for each city

It is possible to focus the right marketing strategy based on the type of customers more present in a specific city



BUSINESS ACTIONS

Clustering and RFM analysis can be used to identify the right type of customer and thus identify which marketing strategy can be done to increase sales and induce buying behaviour in the consumer, trying to tailor one's strategy as much as possible according to the type of customer. RFM analysis or segmentation based on clustering can be used in combination or can be tested to find out which of the proposed solutions is best.

TO BUILD THE
RIGHT
RELATIONSHIPS
WITH THE RIGHT
CUSTOMERS



Strangers

Do not spend too much effort on them

Bernacles

Try to increase their profitability by making offers and finding products they might like.

Old friends

Making them great offers with high discounts.

Promising

Try to build a close relationship, with loyalty programmes, offers on the most purchased products and give them attention with customer care.

Loyals

Increasing consumer loyalty by providing them with a good consumer service and customized offers

True friends

Building an ongoing relationship, with loyalty programmes, rewards and offers reserved for them.

New customers

Great discounts, vouchers and recommend the right products to them.

Identified by

 Clustering

 RFM

 Both

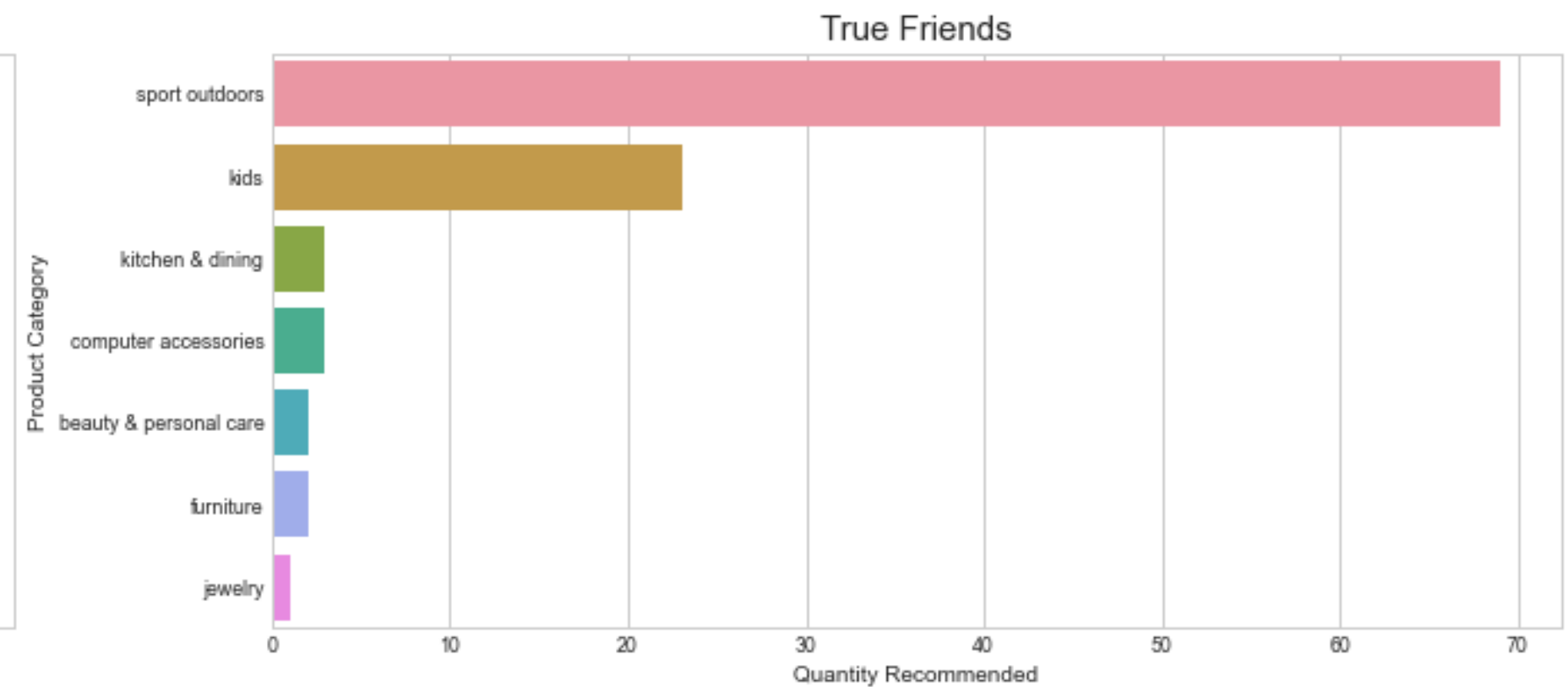
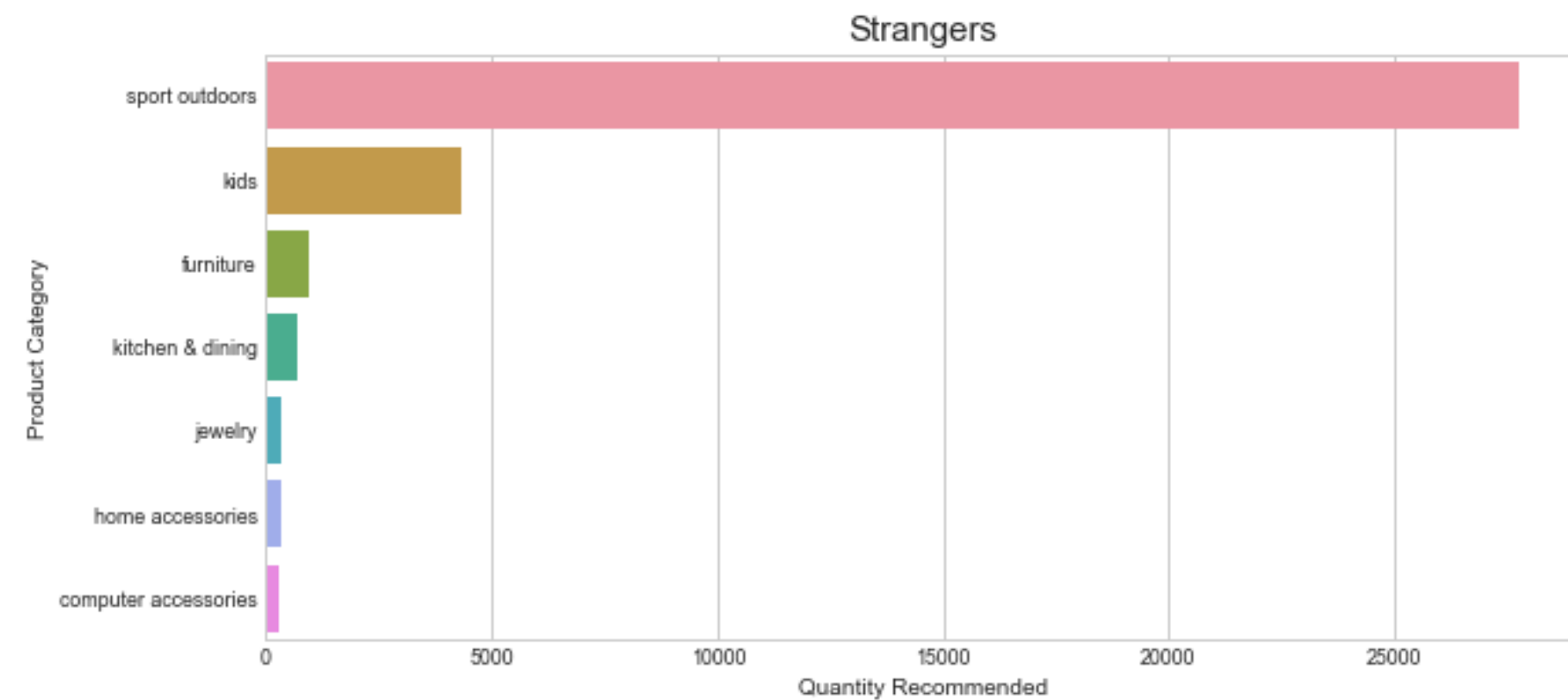
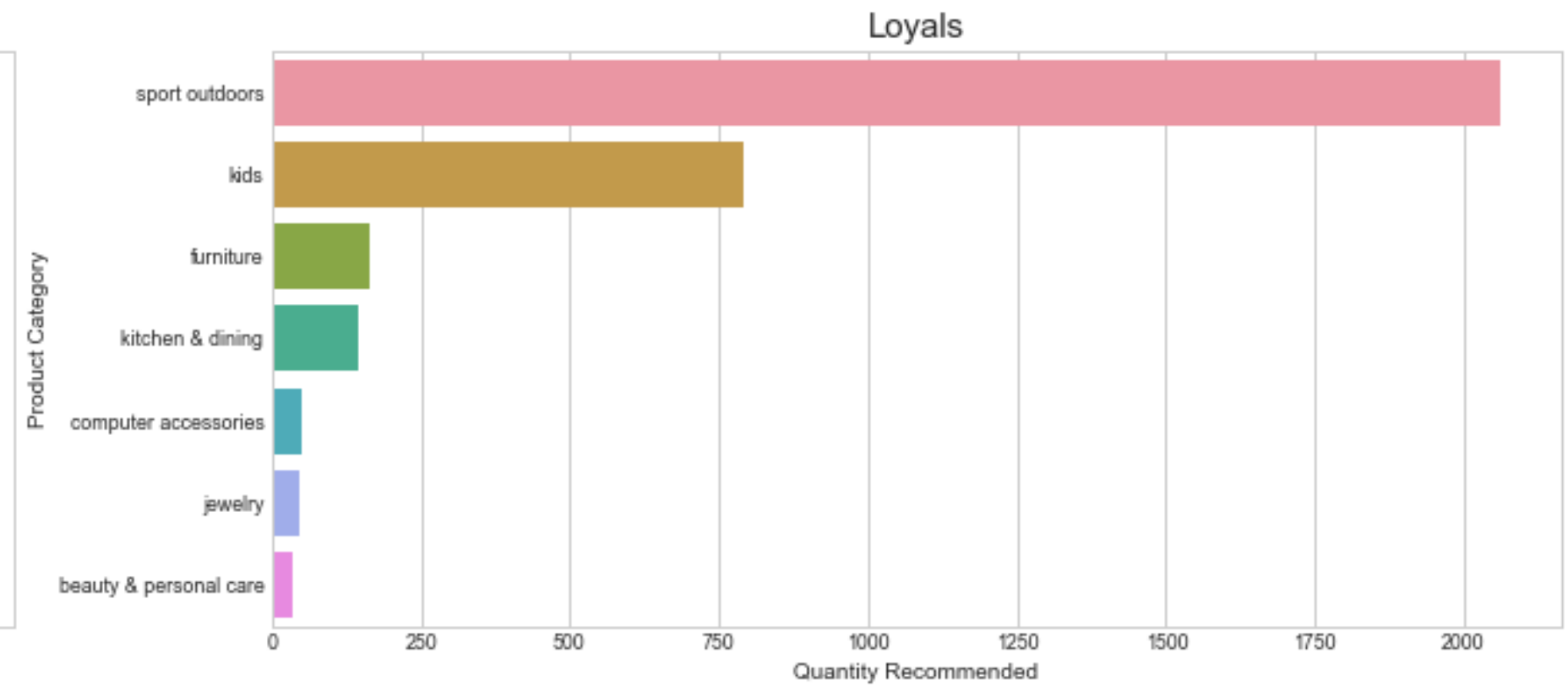
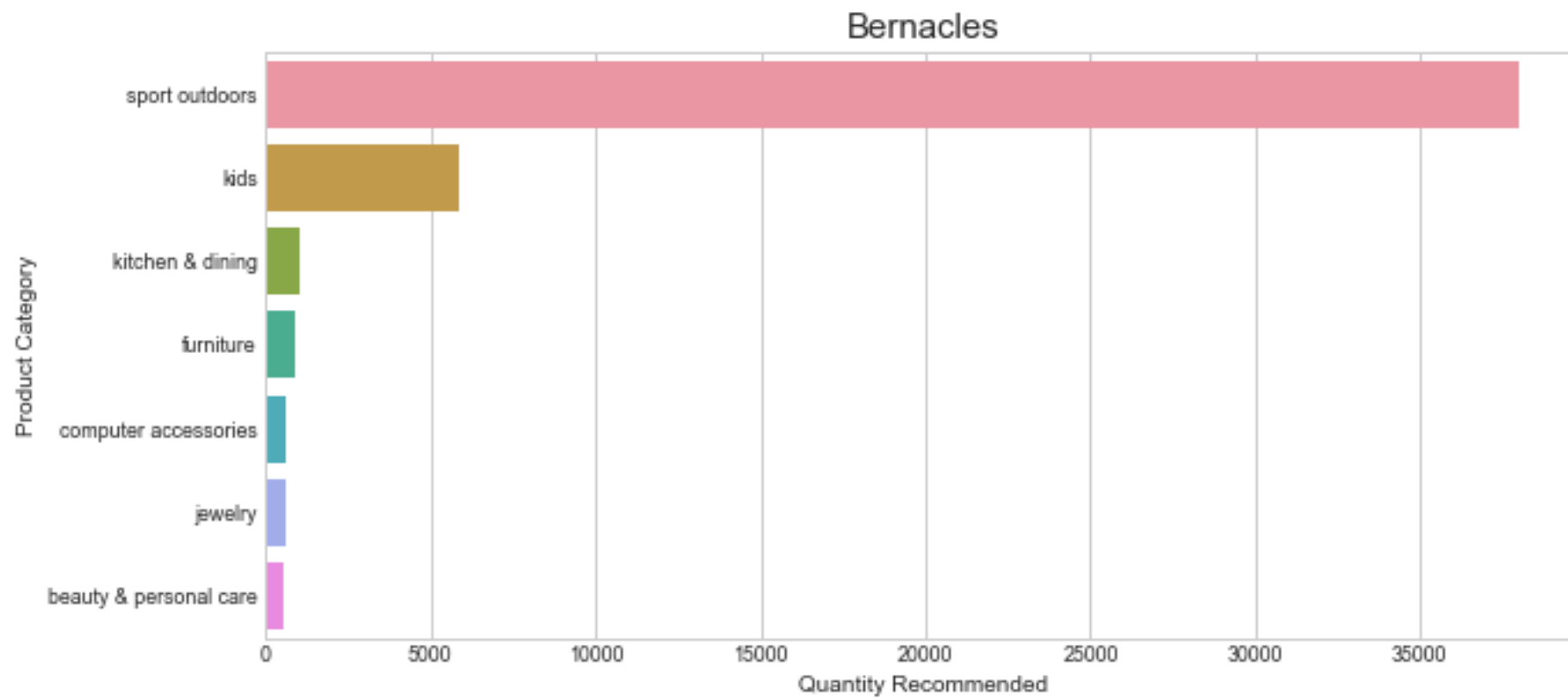
RECOMMENDATION SYSTEM

DEVELOPING THE
ABILITY TO
UNDERSTAND
CUSTOMER
PREFERENCES

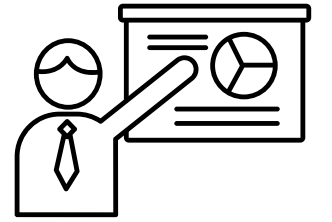
The recommendation system was developed to find the best product to recommend to each customer in order to induce purchasing behaviour.



Most recommended categories by customer segment



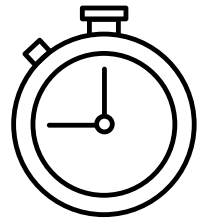
BUSINESS ACTIONS



IT IS POSSIBLE TO SHOW THE RANKING OF PRODUCTS RECOMMENDED, AND THEREFORE POTENTIALLY PREFERRED BY EACH CONSUMER. THE SYSTEM IS OPTIMISED TO RECOMMEND 5 PRODUCTS, BUT CAN BE USED TO CREATE A POTENTIALLY INFINITE LIST OF RECOMMENDED PRODUCTS FOR EACH CONSUMER.

Thanks to this solution it is possible to automatically advice a specific product or product category to a given segment of customer referring both to individual consumer behaviour and comparing the purchasing behaviour of consumers in the same class

DELIVERY TIME PREDICTION



IS IT POSSIBLE TO
OVERPERFORM THE
PREDICTION GIVEN?

Average shipping cost per region of origin

It is possible see for which city the shipping cost is on average higher



PLAN OF ACTION

IN ORDER TO PREDICT THE DAY
OF DELIVERY

Creation of the dataset

Identify the error of the dataset

In order to compare our prediction, we calculated the delivery time error of the company estimation

Train the model

We train three different Machine Learning and Artificial Intelligence models: XGB, ANN, LGBM

Error comparison

In order to identify the best model we plotted each error

LGBM

LightGBM is called “Light” because of its computation power and giving results faster. It takes less memory to run and is able to deal with large amounts of data.

ANN

Exploiting the power of Artificial Intelligence with Artificial Neural Network can boost your business thanks to its reliability and computational power

XGB

The list of benefits and attributes of XGBoost is extensive, and includes the following:

- Usage on a wide range of applications, including solving problems in regression, classification, ranking, and user-defined prediction challenges
- A library that’s highly portable and currently runs on OS X, Windows, and Linux platforms
- Cloud integration that supports AWS, Azure, Yarn clusters, and other ecosystems

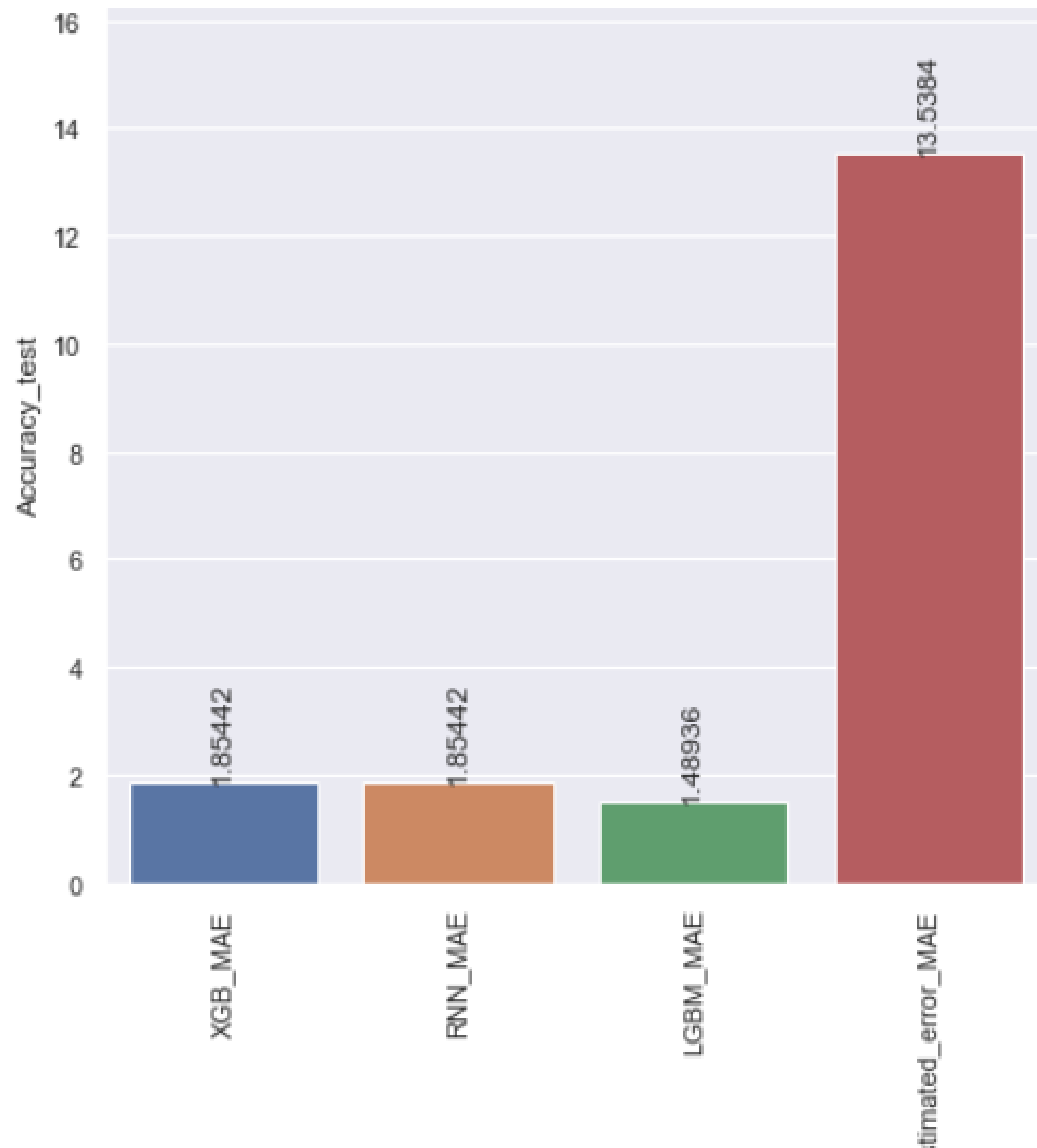
Error:
1.48 DAY (Mean absolute error)

Error:
1.85 DAY (Mean absolute error)

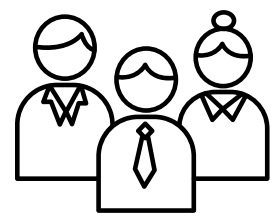
Error:
1.84 DAY (Mean absolute error)

Model comparison

The best model is the LGBM, comparing the company prediction our forecasting is better by 13x



Having more accurate delivery forecasts helps in having a better purchasing experience, showing your customers that the company is reliable



THE TEAM



Miro Confalone



Enrico Grandi



Cosimo Poccianti



Chi Quy Loc