

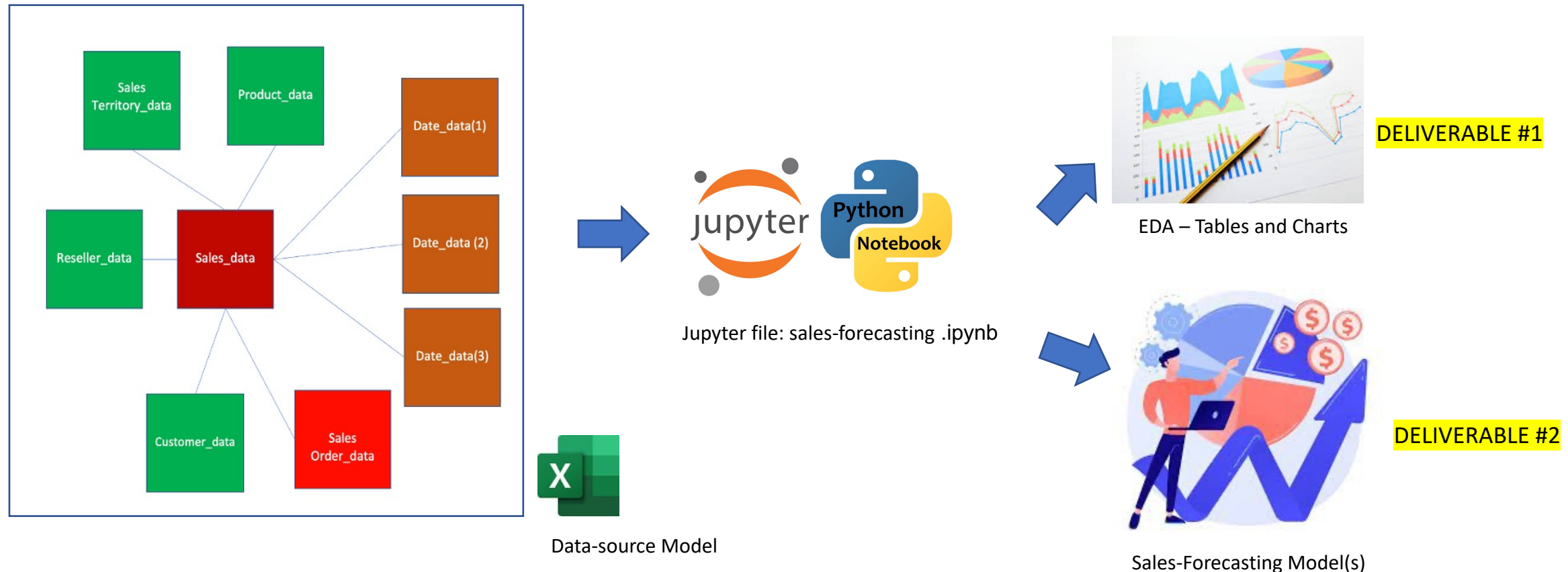


MABJ Corp is developing an analytical application that will support Net Sales forecasting. The company has 2 types of clients (resellers & final customers), which use B2B/B2C platforms to buy products.

A relational database was developed to support the sales transactions and a sample of its main tables was extracted to be used as the source of your insights (deliverable #1) and Machine Learning development (deliverable #2).

You were selected to delivery both **(#1 and #2)**. Python is mandatory to be used.

The data-sample [MABJ-Customer-Sales.xlsx](#) contains the data and the analytical solution to be developed.





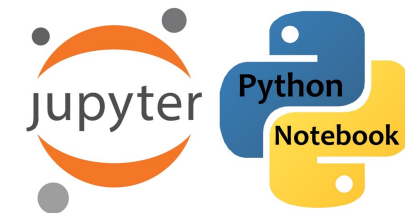
EXAM : GROUP TASKS (3 TEAM MEMBER MAXIMUM)

< Add your full names and student id's to the code >>

< All team members must submit the task >>

DELIVERABLE #1

- Your code must be clear and organized.
- EDA, Tables, Charts (histograms, box-plots, bar charts, ...) and your notes/insights based on data.
- The quality of your analysis will be taken into account to give you a fair grade.
- After develop all analysis and insights, a result must be presented to the board of the company answering this question ... Is it worth it to continue to work with both types of customers ? if yes you must justify and if not, which one is a better option? You must justify in any case.



EXPLORATORY DATA ANALYSIS

Step 1 - Import Dataset and study the dataset carefully.

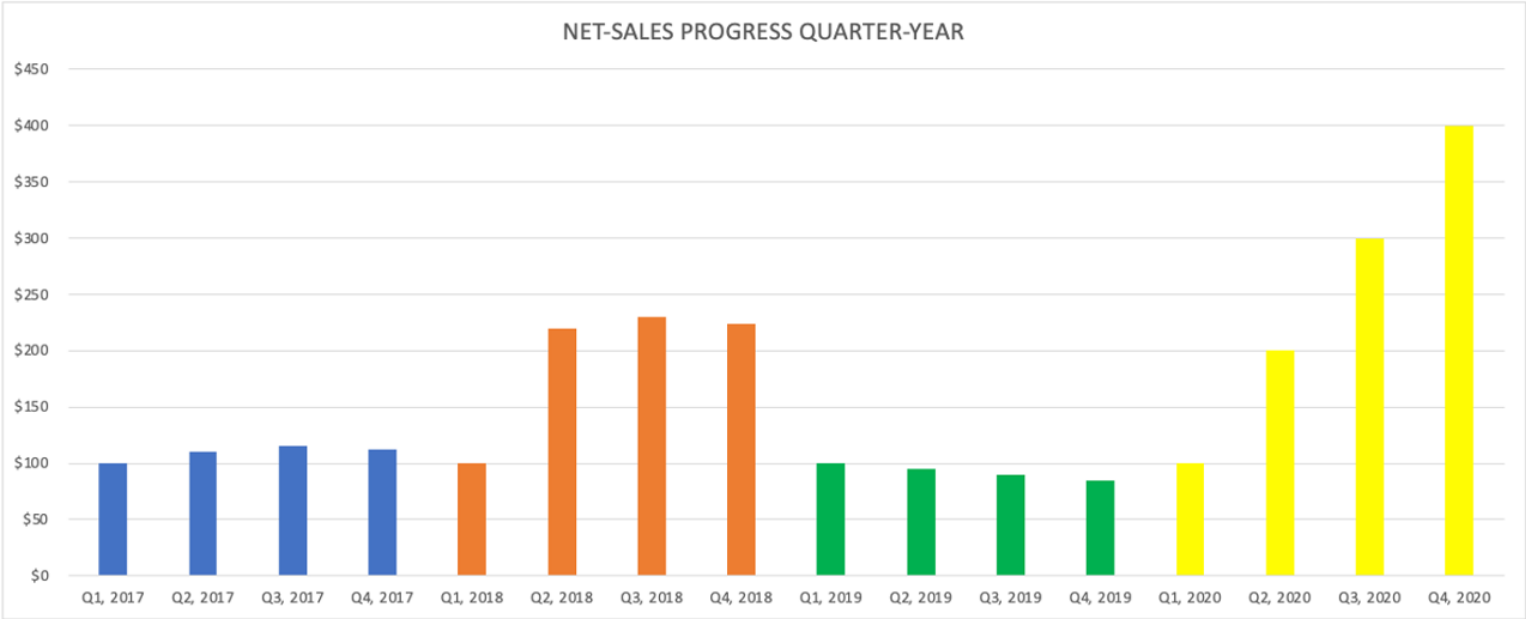
Step 2 - Add a new feature to your dataset. Net-sales per transaction. << Net Sales = Gross Sales – Total cost >>

Step 3 – Analysis (at least these 3 data analysis). Develop more is recommended.



ANALYSIS (#1)

QUARTE-YEAR	NET-SALES
Q1, 2017	\$100
Q2, 2017	\$110
Q3, 2017	\$115
Q4, 2017	\$112
Q1, 2018	\$100
Q2, 2018	\$220
Q3, 2018	\$230
Q4, 2018	\$224
Q1, 2019	\$100
Q2, 2019	\$95
Q3, 2019	\$90
Q4, 2019	\$85
Q1, 2020	\$100
Q2, 2020	\$200
Q3, 2020	\$300
Q4, 2020	\$400



DOES NOT REPRESENT THE DATASET, ONLY TO ILLUSTRATE
INSIGHTS ABOUT NET SALES QUARTER

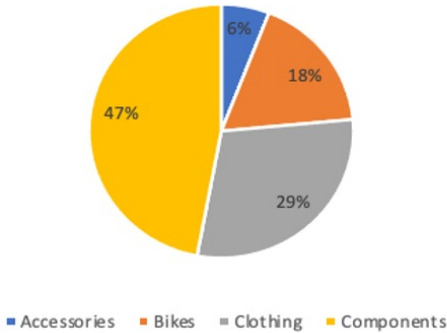
Not only chart(s), you must add your analysis about the data

ANALYSIS (#2)

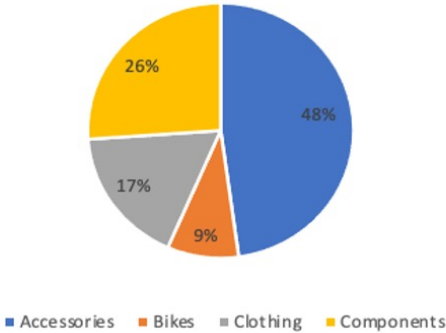
YEAR	CATEOGRY	TOTAL-COST (%)
2017	Accessories	6%
	Bikes	18%
	Clothing	29%
	Components	47%
2018	Accessories	48%
	Bikes	9%
	Clothing	17%
	Components	26%
2019	Accessories	32%
	Bikes	12%
	Clothing	28%
	Components	28%
2020	Accessories	28%
	Bikes	28%
	Clothing	31%
	Components	13%

Note: Total-Cost (Order Quantity * Product Cost)

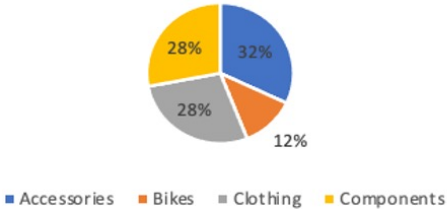
2017 (COST-DISTRIBUTION)



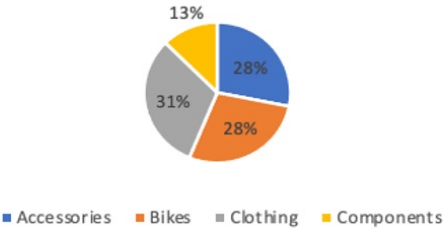
2018 (COST-DISTRIBUTION)



2019 (COST-DISTRIBUTION)



2020 (COST-DISTRIBUTION)



DOES NOT REPRESENT THE DATASET, ONLY TO ILUSTRATE INSIGHTS ABOUT ANALYSIS

Not only chart(s), you must add your analysis about the data

ANALYSIS (#3)

YEAR	TOP3 NET SALES (RESELLER)	TOP3 NET SALES (CUSTOMER)
2017	TOP-1	TOP-1
	TOP-2	TOP-2
	TOP3	TOP3
2018	TOP-1	TOP-1
	TOP-2	TOP-2
	TOP3	TOP3
2019	TOP-1	TOP-1
	TOP-2	TOP-2
	TOP3	TOP3
2020	TOP-1	TOP-1
	TOP-2	TOP-2
	TOP3	TOP3

Not only chart(s), you must add your analysis about the data

DELIVERABLE #2



Sales-Forecasting Model(s)



- The company would like to predict **Net Sales** regarding its best selected type of customer (choose the best one).
- Show your model, training dataset features, dataset split strategy and outcomes of your model given fake entries.
- Your code must be clear and organized.