## TASKS

The tasks below will be completed before the project presentation. Write your name next to any task.

#### 1. Data Cleaning & Exploratory Data Analysis

* Import Modules, Load Data & Data Review
* Follow the Steps Below  
  *i. Take a look at relationships between InvoiceNo, Quantity and UnitPrice columns.  
  ii. What does the letter "C" in the invoiceno column mean?  
  iii. Handling Missing Values  
  iv. Clean the Data from the Noise and Missing Values  
  v. Explore the Orders  
  vi. Explore Customers by Country  
  vii. Explore the UK Market*

#### 2. RFM Analysis

* Follow the steps below  
  *i. Import Libraries  
  ii. Review "df\_uk" DataFrame (the df\_uk what you create at the end of the Task 1)  
  iii. Calculate Recency  
  iv. Calculate Frequency  
  v. Calculate Monetary Values  
  vi. Create RFM Table*

#### 3. Customer Segmentation with RFM Scores

* Calculate RFM Scoring  
  *i. Creating the RFM Segmentation Table*
* Plot RFM Segments

#### 4. Applying K-Means Clustering

* Data Pre-Processing and Exploring  
  *i. Define and Plot Feature Correlations  
  ii. Visualize Feature Distributions  
  iii. Data Normalization*
* K-Means Implementation  
  *i. Define Optimal Cluster Number (K) by using "Elbow Method" and "Silhouette Analysis"  
  ii. Visualize the Clusters  
  iii. Assign the label  
  iv. Conclusion*

#### 5. Create Cohort and Conduct Cohort Analysis

* Future Engineering  
  *i. Extract the Month of the Purchase  
  ii. Calculating time offset in Months i.e. Cohort Index*
* Create 1st Cohort: User Number & Retention Rate  
  *i. Pivot Cohort and Cohort Retention  
  ii. Visualize analysis of cohort 1 using seaborn and matplotlib*
* Create 2nd Cohort: Average Quantity Sold  
  *i. Pivot Cohort and Cohort Retention  
  ii. Visualize analysis of cohort 2 using seaborn and matplotlib*
* Create 3rd Cohort: Average Sales  
  *i. Pivot Cohort and Cohort Retention  
  ii. Visualize analysis of cohort 3 using seaborn and matplotlib*
* **Note: There may be sub-tasks associated with each task, you will see them in order during the course of the work.**

Each person should start doing the task immediately. Noone should wait one another. Only EDA can be awaited to finish. Others are independent.

When someone finishes his own task should share the insights and details to improve the solution. He can also try helping others to finish.

## READING TASKS

The reading tasks are free to follow for each colleague. They help doing the work in a better way.

### RFM Analysis:

<https://www.datacamp.com/community/tutorials/introduction-customer-segmentation-python?utm_source=adwords_ppc&utm_medium=cpc&utm_campaignid=16079694435&utm_adgroupid=&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=x&utm_adpostion=&utm_creative=&utm_targetid=&utm_loc_interest_ms=&utm_loc_physical_ms=1012783&gclid=Cj0KCQjw6J-SBhCrARIsAH0yMZhg-ZXbQdRHLOLDZPQmf2VMFxe6g9wR_WVmKvpdJ9u4Ec2vBE0V6Q4aAre1EALw_wcB>

### Data Visualization Sample Capstone Project:

<https://drive.google.com/file/d/1LC_F5SblLCLHoo3m0GvcdyKlaHJAEQQd/view?usp=sharing>

### Similar Project: Sentiment Classification

<https://www.kaggle.com/code/kadirduran/nlp-sentiment-classification-with-ml-and-dl-models/notebook>

### Plotting in python using Seaborn, Matplotlib and Plotly (article):

<https://dylancastillo.co/how-to-plot-with-python-popular-graphs-using-pandas-matplotlib-seaborn-and-plotly-express/>

### Cohort Analysis

<https://clevertap.com/blog/cohort-analysis/>

<https://www.appcues.com/blog/cohort-analysis>

### K-Means Segmentation Sample Project

<https://drive.google.com/file/d/1y98bPKxCvMYqsElnMjxN6XStSYyc_63F/view?usp=sharing>