Capstone -2 : Unsupervised Learning Customer Segmentation

Approach:

EDA Part:

1. Analyzed data and checked for Missing values
   Analysis on some Important fields in Data:
2. Found missing values in CustomerID – which is main field – so we can apply Imputation like Mean, Median but if we apply that then it will be Bias for 1 particular customer , hence dropped the rows which has null customerID.
3. Found 4372 Unique Customers.
4. Analyzed InvoiceNo field where few Invoice are in Cancel state , so checked whether XYZ = 'C'XYZ   - if I found that I could minus the billing amount or remove that Invoice may impact analysis , but I count found some matched there
5. Found 8905 cancelled order
6. Found Billing Amount by Multiplying Unit Price By Quantity
7. Dropped Unwanted columns and prepared final data with InvoiceNo, InvoiceDate [Did some analysis on Date as well], Bills/Revenue & CustomerID

RFM:

- Applied directly Recency , frequency & Monetory on Final data
- Calculated RFM score by taking Lowest Recency, Highest Frequency & Highest Monetory
- Lowest the score = Top most customer's are segregated

Attached Python file with detailed analysis.