**Hoàng Nguyễn Minh Du - K184060779 - K18406C**

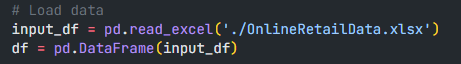
**Object:**

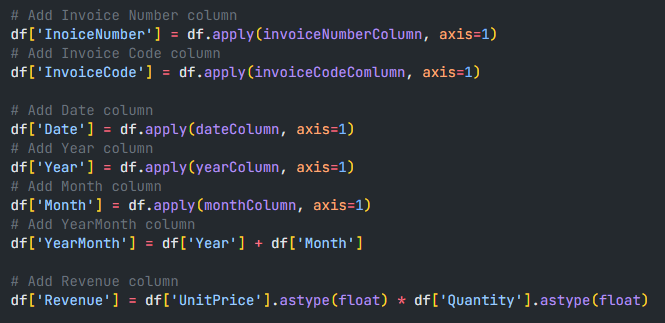
Create a list of target customer with good insights

**Solve:**

**Date preparing (File DataPrep.py)**

1. Read and add needed columns for analyse

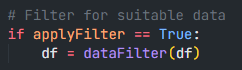




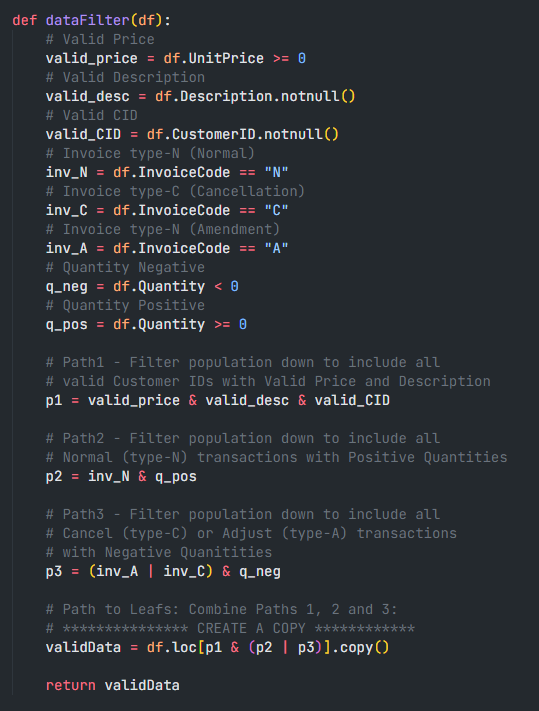
* 1. Return suitable result for columns



1. Conditions to apply filter

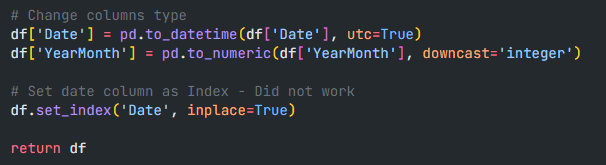


* 1. Filter detail



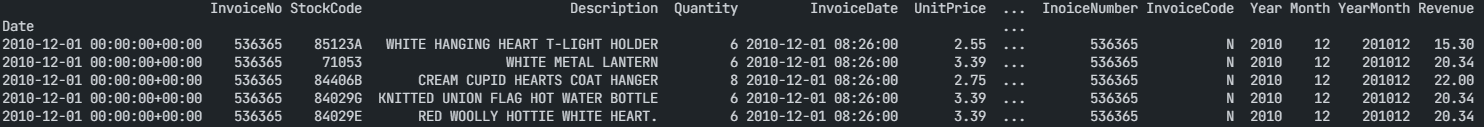
Data such as price, description, ID, quantity is not allowed to be empty or smaller then 0

1. Validate and return result



Change column type and set column index for further step

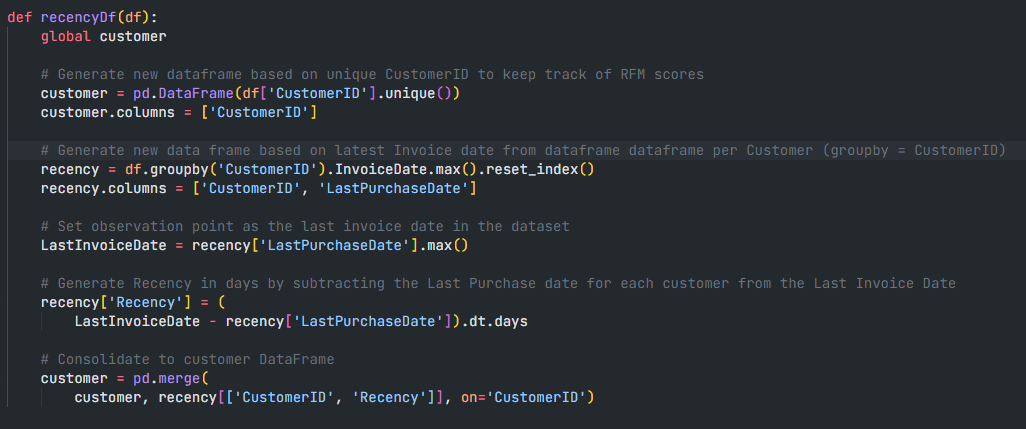
Result: First 5 rows



**RFM Analysis (file RFMAnalysis.py)**

Create a global dataframe called customer to return

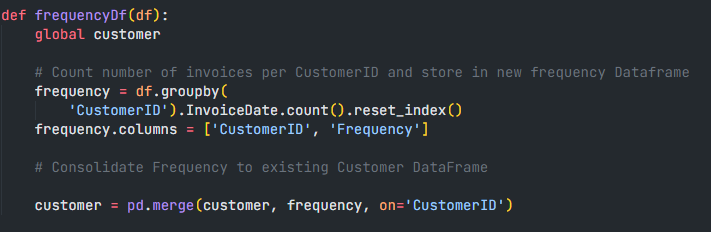
1. Recency



Add CustomerID column from input dataframe

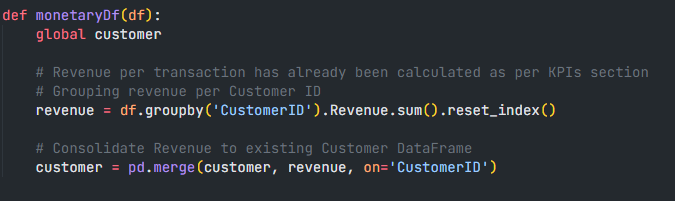
Add Rencency column base on latest invoice date and last purchase date

1. Frequency



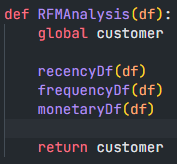
Add Frequency column base on the number of invoice date came from each customer

1. Monetary

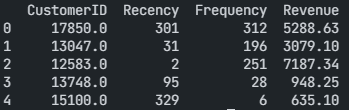


Add the total amount of money came from customer

1. Return result



Result: First 5 rows



**Analyse (File Analysis.py)**

1. Giving author information and call main analysis function



1. Read and validate input data



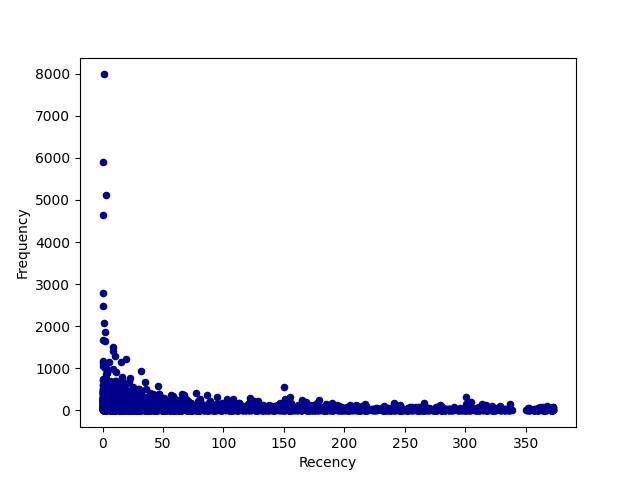
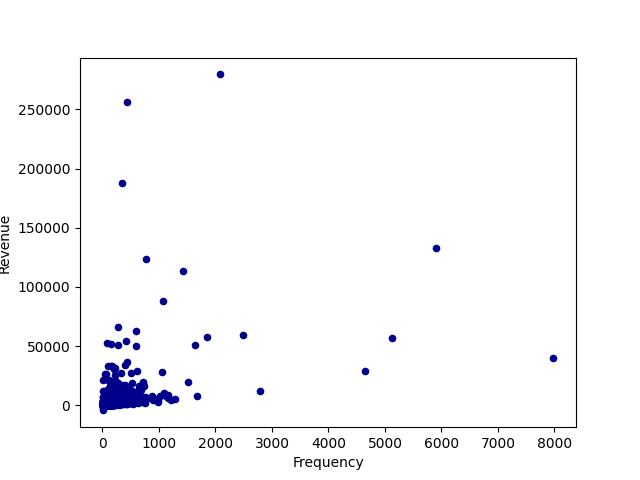
1. RFM analysis

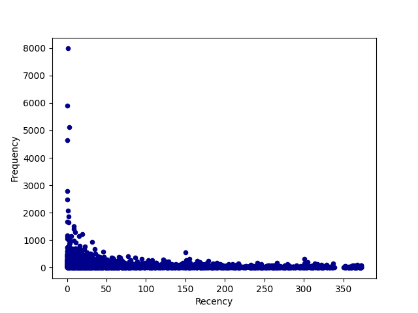


1. Save RFM analysis image



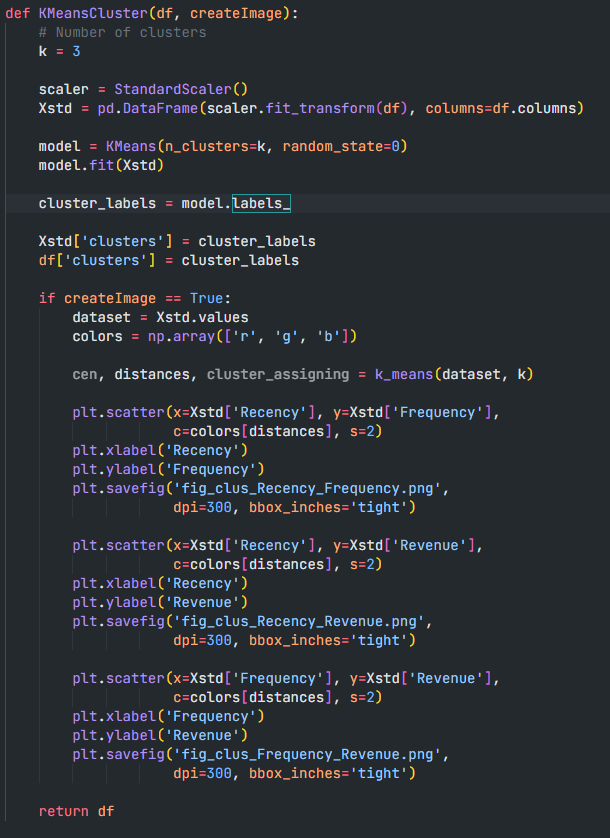
Result:



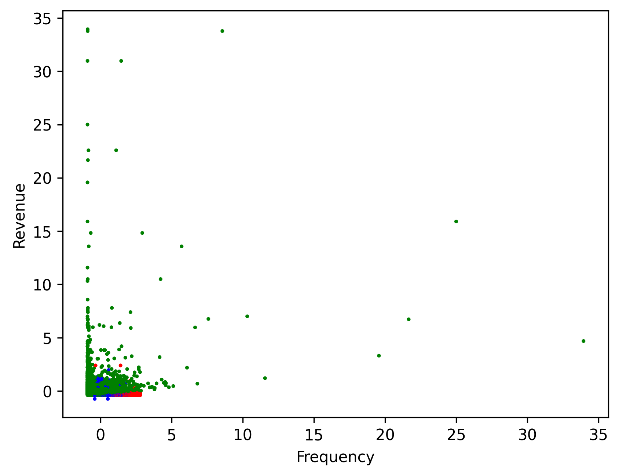
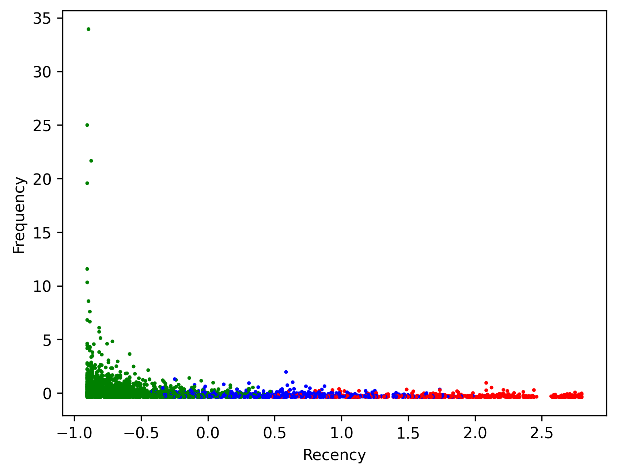


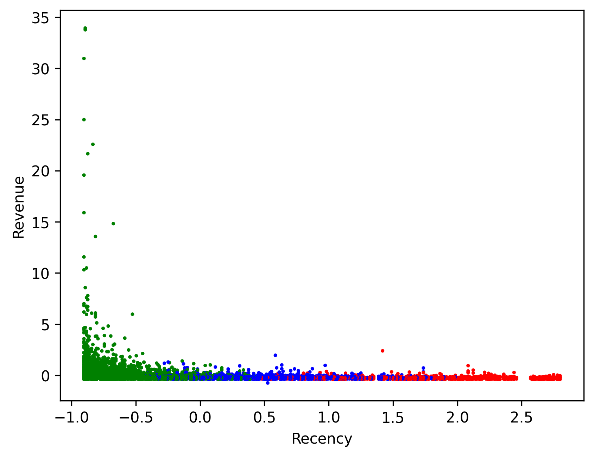
1. Using KMeans and save analysis image





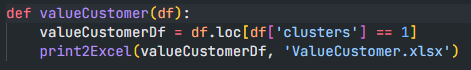
* Create a new dataframe with input dataframe and scaler
* Using KMeans with k = 3
* Add cluster column
* Save image
* Create dataset from above dataframe
* Create colors array
* Calculate distance from centroids
* Add image detail and save figure
* Result:



1. Create a list of customers with good insights





Result is in file ValueCustomer.xlsx

**Conclusion:**

Created successfully a list of good customer