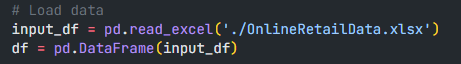
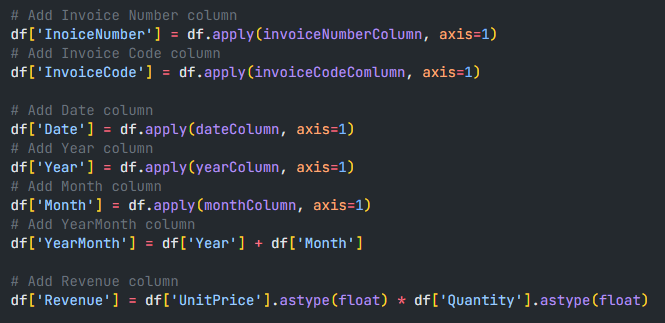
**Hoàng Nguyễn Minh Du - K184060779 - K18406C**

**Chuẩn bị dữ liệu (File DataPrep.py)**

1. Đọc và thêm các cột cần thiết để phân tích

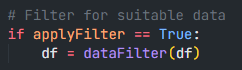




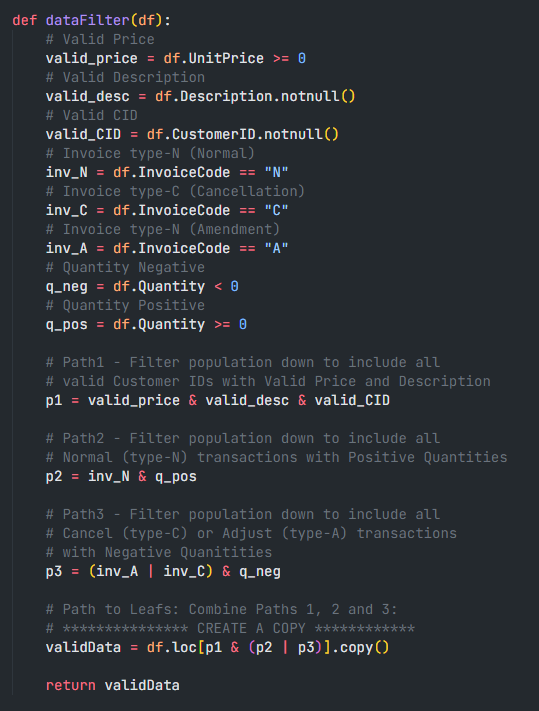
* 1. Điều kiện để cho các dữ liệu của cột được tạo ra



1. Điều kiện để lọc các đơn hàng phù hợp

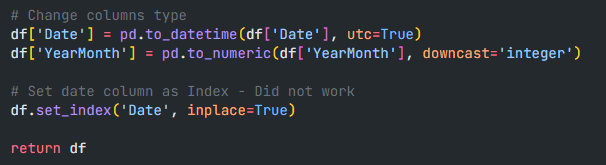


* 1. Chi tiết điều kiện lọc



Các dữ liệu như giá, tên hàng, mã khách hàng, mã hóa đơn, số lượng mua không được bỏ trống hoặc nhỏ hơn không tùy loại

1. Chuẩn hóa và trả kết quả

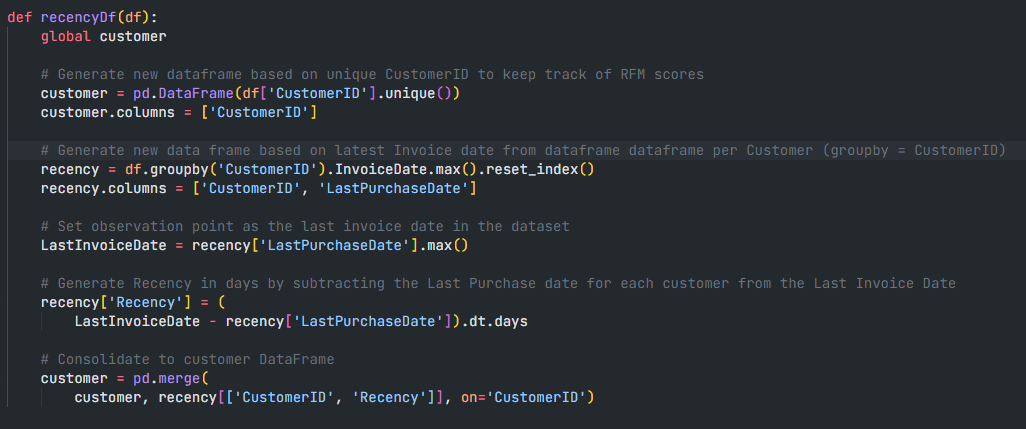


Đổi kiểu dữ liệu để set index cho cột hoặc để có thể so sánh ở các bước sau

**Phân tích RFM (file RFMAnalysis.py)**

* Tạo ra dataframe customer global để trả về phân tích tiếp các bước sau

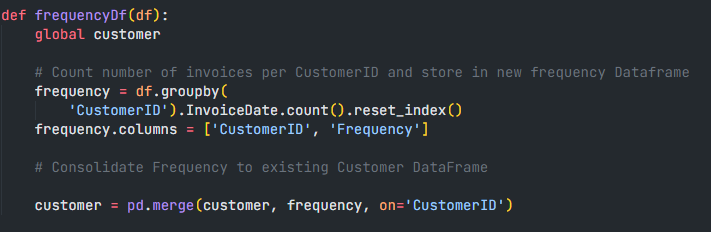
1. Recency



Với dataframe customer:

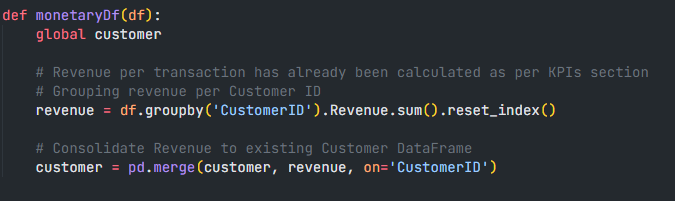
* Thêm cột CustomerID
* Thêm cột Recency dựa vào thời gian của hóa đơn gần nhất và thời gian mua gần nhất

1. Frequency



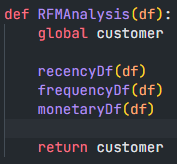
Thêm vào dataframe customer cột frequency dựa vào số lần mua

1. Monetary



Thêm vào dataframe customer cột monetary dựa vòa doanh thu khách hàng mang lại

1. Trả kết quả



**Phân tích (File Analysis.py)**

1. Ghi thông tin và gọi hàm phân tích chính



1. Đọc và chuẩn hóa dữ liệu (chi tiết phía trên)



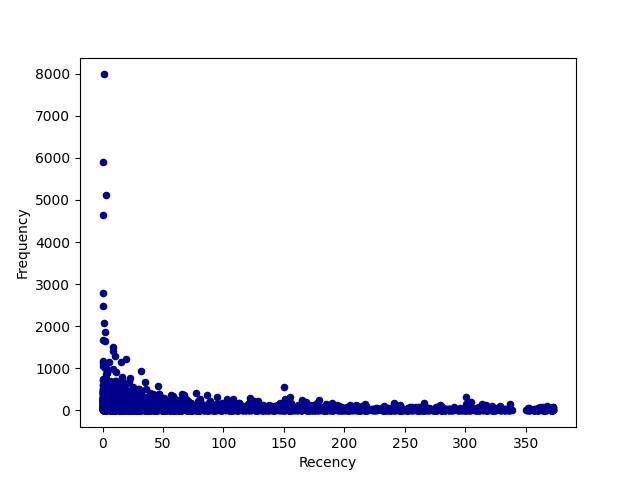
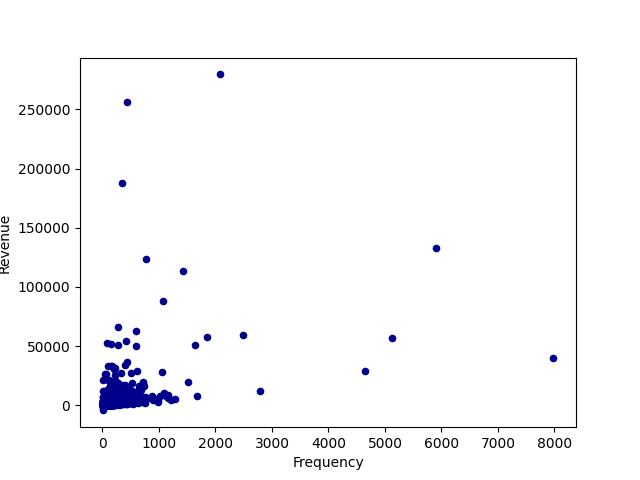
1. Phân tích RFM (chi tiết phía trên)

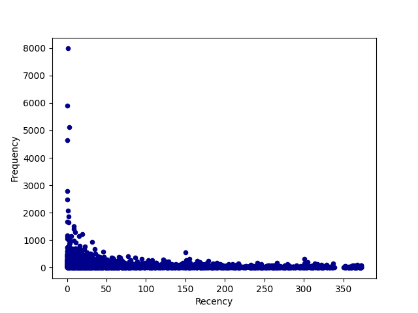


1. Ghi ra file hình kết quả phân tích



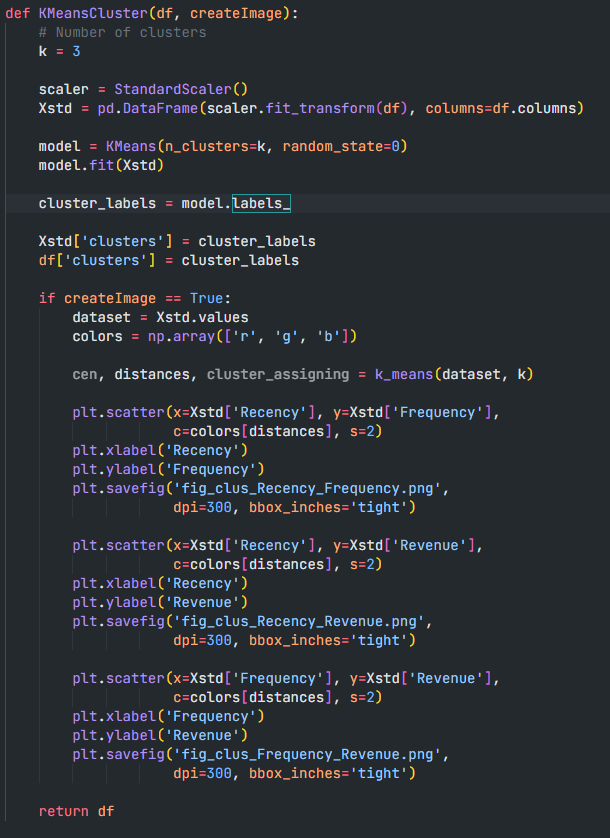
Kết quả:



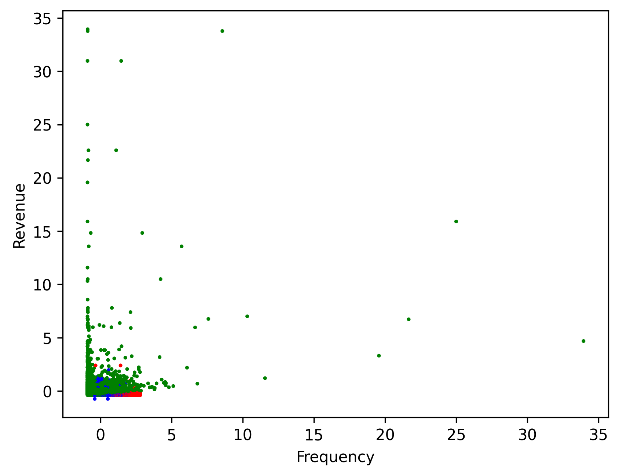
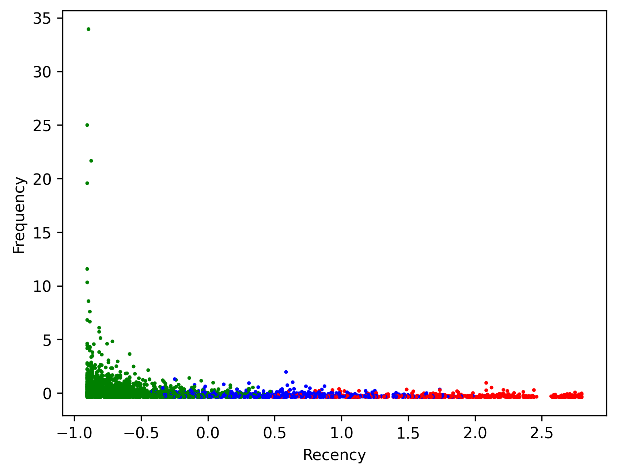


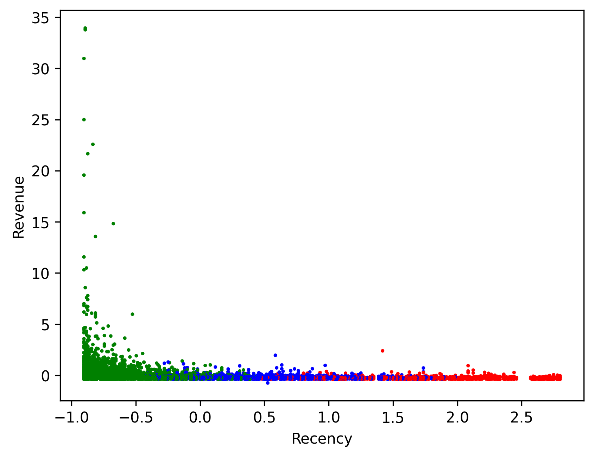
1. Gom nhóm bằng KMeans và in ra kết quả





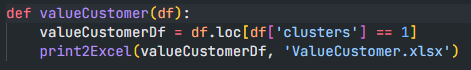
* Tạo mảng
* Tạo 1 dataframe mới từ dataframe đưa vào kết hợp với scaler
* Dùng thuật toán Kmeans để gom nhóm với k = 3
* Thêm cột cluster để phân loại
* Ghi ra hình
* Tạo dataset từ dataframe ở trên
* Tạo mảng màu để phân biệt các nhóm
* Tính khoảng cách so với tâm để chuyển màu cho các điểm
* Thêm các thông tin hình và ghi ra file hình
* Kết quả



1. In ra danh sách các khách hàng tốt





Kết quả ở file excel ValueCustomer.xlsx