📊 Retail Sales Analysis – Project Report

Name: Neha Raut
Role: Data Analyst
Tools Used: MySQL (MySQL Workbench), Python (Jupyter Notebook), Google Sheets
Project: Retail Sales Analysis
Date:31/12/2025

🧾 1. Introduction

This project focuses on analyzing retail sales data to uncover insights related to sales performance, product trends, and regional contribution. The goal is to transform raw transactional data into meaningful business insights using **MySQL for querying**, **Python (Jupyter Notebook) for cleaning and analysis**, and **Google Sheets for pivot-table based reporting and charts**.

📌 2. Project Overview

Domain: Retail / Sales Analytics
Description: The project analyzes historical retail sales data to identify trends, top-performing categories and products, and region-wise performance. The final output is a structured Google Sheets report with pivot tables and charts for easy interpretation.

❓ 3. Problem Statement

Retail businesses generate large volumes of sales data but often lack clear visibility into: - Which categories and products drive most revenue - Which regions perform better or worse - How sales change over time (monthly / yearly)

The business needs a simple analytical report to monitor performance and support data-driven decisions.

🎯 4. Objectives

- Analyze overall sales and quantity trends
- Identify top and low performing categories and products
- Compare performance across regions
- Create pivot-table based reports and charts in Google Sheets

- Demonstrate end-to-end data analysis workflow using SQL and Python

---

🗃️ 5. Data Description

- **Source:** Retail sales dataset (CSV)
- **Type:** Transactional sales data
- **Main Columns:** Order Date, Product, Category, Sub-Category, Sales, Quantity, Profit, Region, State
- **Data Issues:** Missing values, inconsistent category names, date format issues (handled during cleaning)

🛠️ 6. Tools & Technologies Used

- **MySQL (MySQL Workbench):** Data storage, SQL queries, aggregations
- **Python (Jupyter Notebook):** Data cleaning, preprocessing, exploratory analysis (pandas, numpy, matplotlib)
- **Google Sheets:** Pivot tables and charts for reporting and visualization

---

🔄 7. Methodology / Workflow

1. Collected retail sales data (CSV)
2. Loaded data into MySQL database
3. Performed SQL queries for initial analysis (aggregations, grouping)
4. Exported data to Jupyter Notebook
5. Cleaned and prepared data using Python
6. Performed exploratory data analysis
7. Created pivot tables and charts in Google Sheets

⚙️ 8. Implementation Details

8.1 MySQL (MySQL Workbench)

- Created table for sales data
- Performed queries such as:
    - Total sales by category and region
    - Monthly and yearly sales summary
    - Top products by sales
- Used SQL concepts: SELECT, WHERE, GROUP BY, ORDER BY, aggregate functions

8.2 Jupyter Notebook (Python)

- Data cleaning:
  - Handled missing values
  - Converted date columns to proper format
  - Standardized category and region names
- Feature engineering:
  - Created Year and Month columns from Order Date
- Analysis performed:
  - Monthly sales trend
  - Category-wise and region-wise performance
- Libraries used: pandas, numpy, matplotlib

8.3 Google Sheets (Pivot Tables & Charts)

- Created multiple **pivot tables** to summarize:
  - Sales by Category and Sub-Category
  - Sales by Region and State
  - Monthly and Yearly Sales Trends
- Built **charts** from pivot tables:
  - Bar charts for category and region comparison
  - Line charts for time-based trends
  - Column charts for top products

📈 9. Results & Insights

- Identified top-performing categories and sub-categories
- Found regions contributing the highest sales
- Discovered clear seasonal trends in monthly sales
- Observed that some categories contribute more to volume but less to revenue
- These insights can help in planning promotions, inventory, and regional strategies

---

📊 10. Visualization (Google Sheets)

The Google Sheets report contains: - **Pivot Table Sheets:** - Sales by Category and Sub-Category - Sales by Region and State - Monthly Sales Trend - **Charts Sheets:** - Category-wise Sales Chart - Region-wise Sales Chart - Monthly Sales Trend Line Chart

These visuals make it easy for non-technical users to understand business performance.

⚠️ 11. Challenges & Solutions

Challenge	Solution
Missing / inconsistent data	Cleaned and standardized using Python
Date format issues	Converted to proper datetime format in Python
Large data summarization	Used SQL and pivot tables for efficient aggregation

---

✅ 12. Conclusion

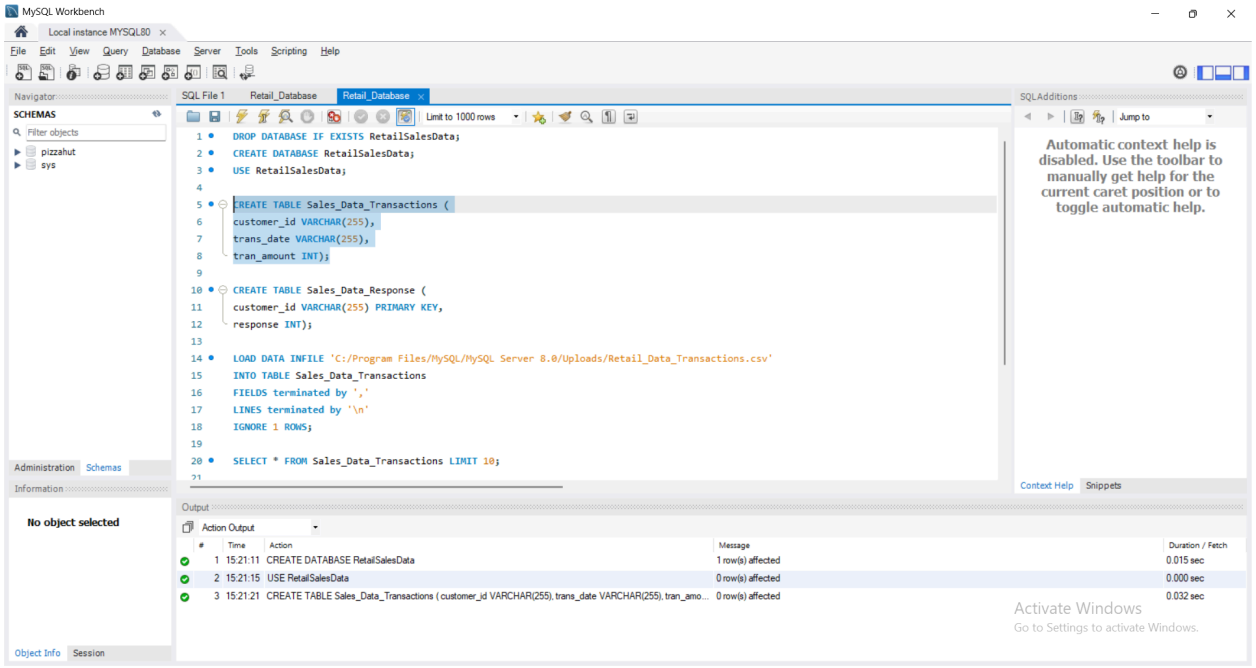
This Retail Sales Analysis project demonstrates a complete data analysis workflow from raw data to business insights. Using MySQL, Python, and Google Sheets, the project delivers a simple yet effective reporting solution that helps understand sales performance and supports data-driven decision making.

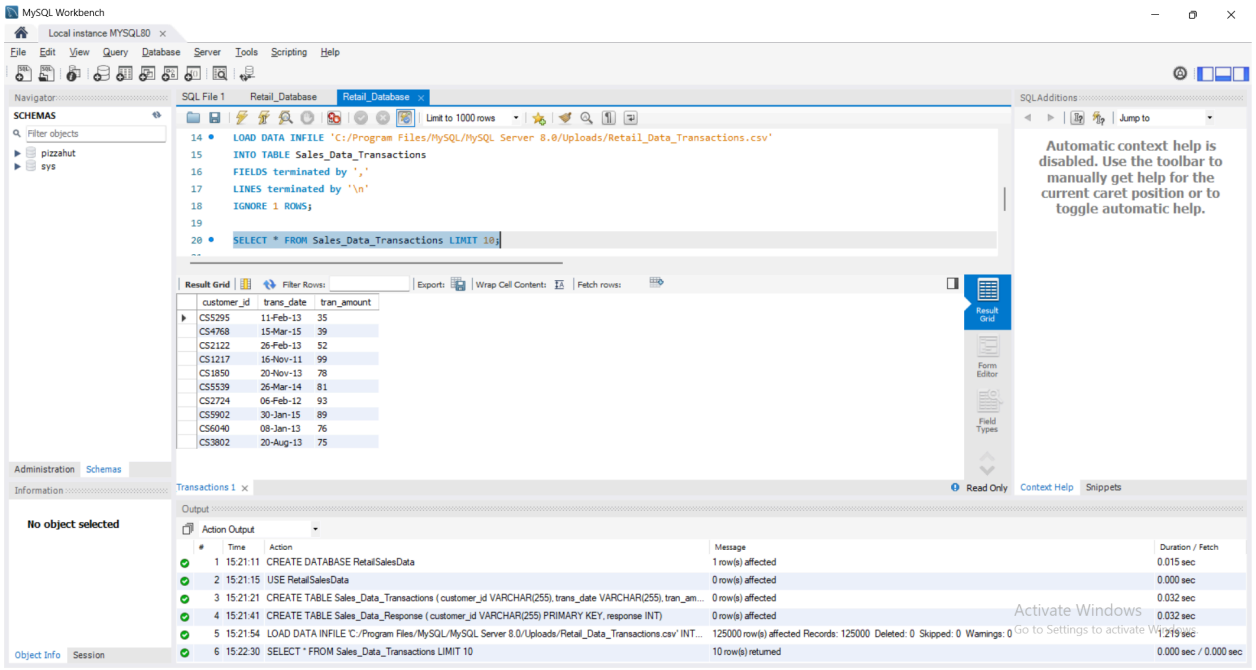
🔮 13. Future Improvements

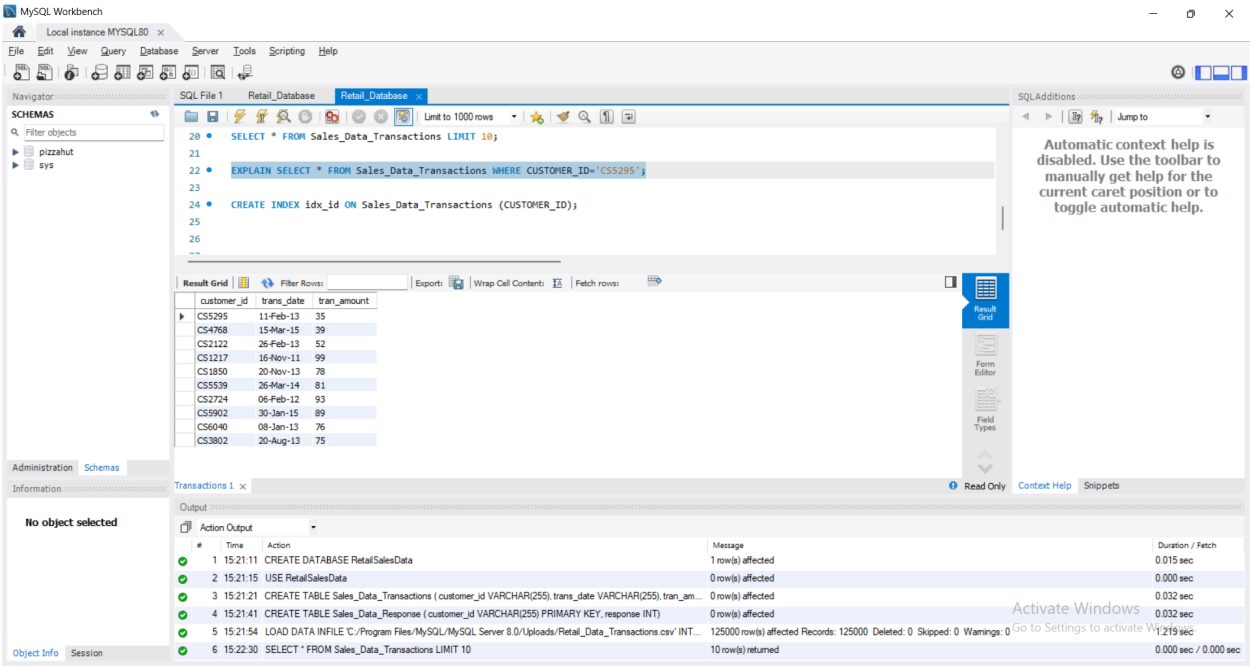
- Add more recent or real-time data
- Add profit and customer-level analysis
- Automate data update process
- Add forecasting using Python

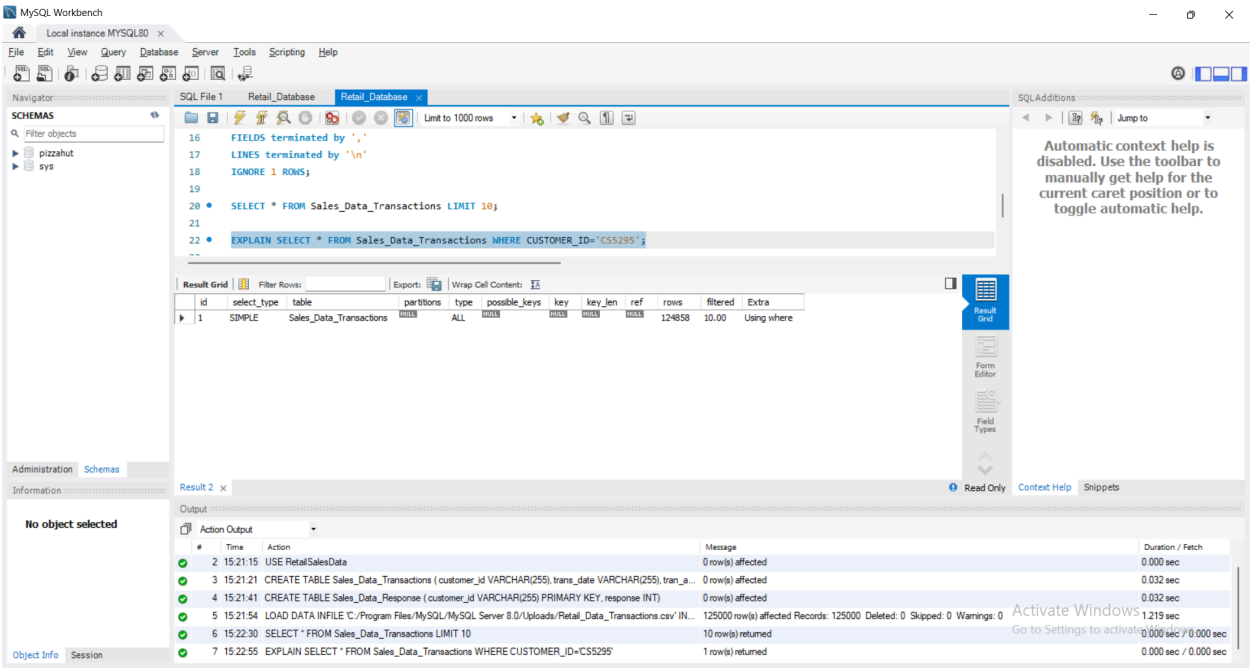
📎 14. Screenshots

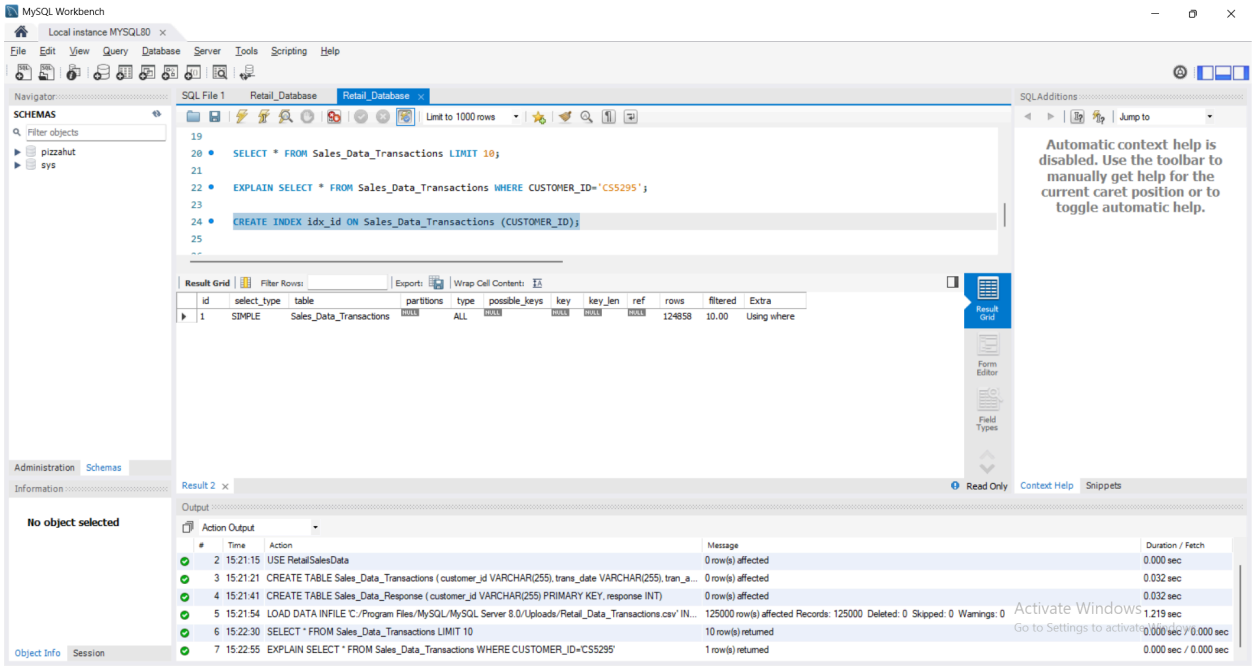
- Jupyter Notebook analysis outputs

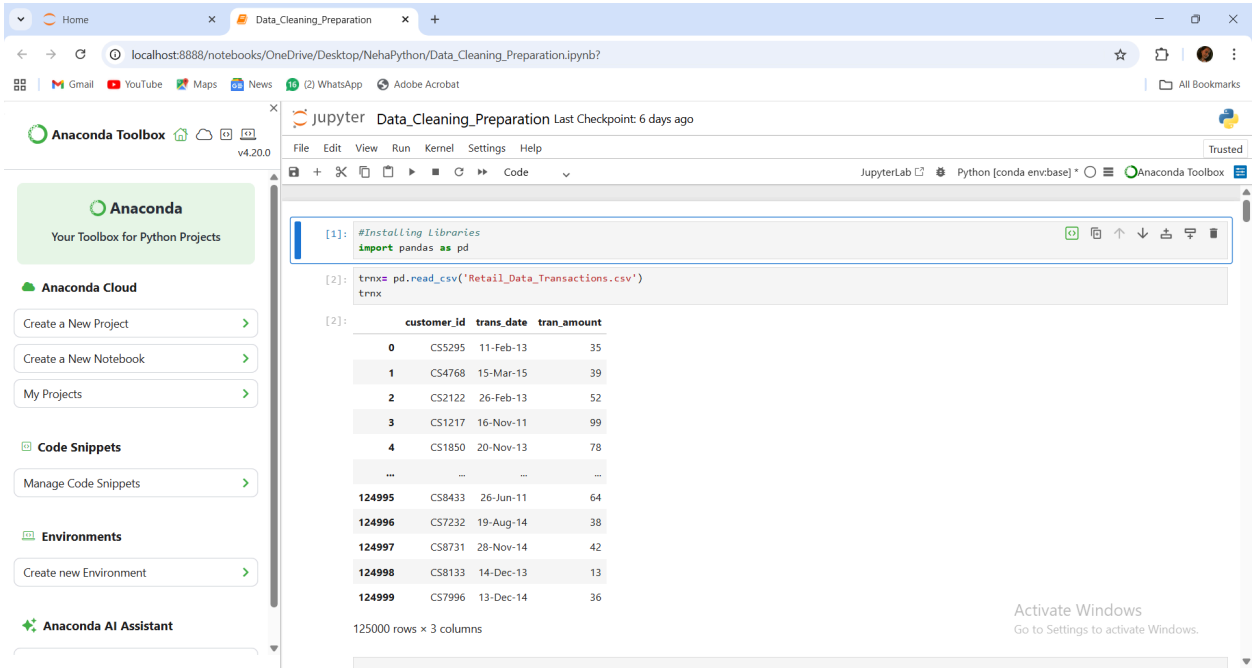

MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator:
SCHEMAS
Filter objects
pizzahut
sys
Administration Schemas
Information
No object selected
Object Info Session
SQL File 1 Retail_Database Retail_Database
Limit to 1000 rows
1 DROP DATABASE IF EXISTS RetailSalesData;
2 CREATE DATABASE RetailSalesData;
3 USE RetailSalesData;
4
5 CREATE TABLE Sales_Data_Transactions (
6 customer_id VARCHAR(255),
7 trans_date VARCHAR(255),
8 tran_amount INT);
9
10 CREATE TABLE Sales_Data_Response (
11 customer_id VARCHAR(255) PRIMARY KEY,
12 response INT);
13
14 LOAD DATA INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/Retail_Data_Transactions.csv'
15 INTO TABLE Sales_Data_Transactions
16 FIELDS terminated by ','
17 LINES terminated by '\n'
18 IGNORE 1 ROWS;
19
20 SELECT * FROM Sales_Data_Transactions LIMIT 10;
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
1 15:21:11 CREATE DATABASE RetailSalesData 1 row(s) affected 0.015 sec
2 15:21:15 USE RetailSalesData 0 row(s) affected 0.000 sec
3 15:21:21 CREATE TABLE Sales_Data_Transactions (customer_id VARCHAR(255), trans_date VARCHAR(255), tran_amo... 0 row(s) affected 0.032 sec
Activate Windows
Go to Settings to activate Windows.

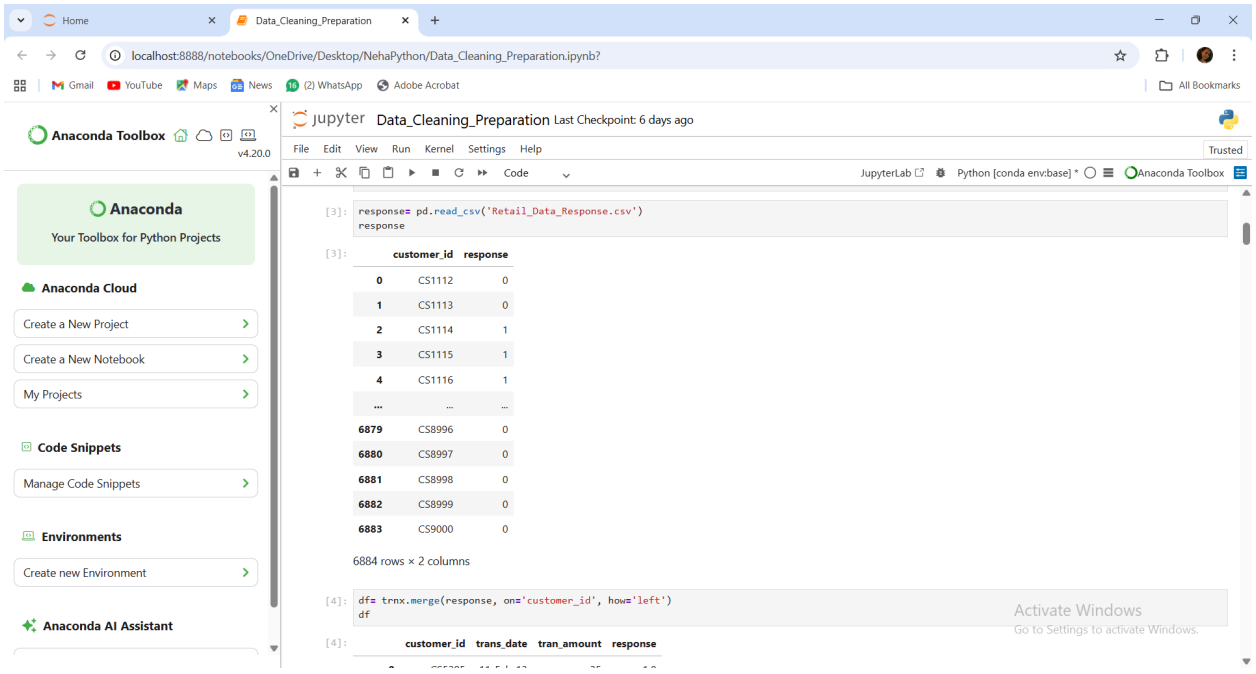

MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator:
SCHEMAS
Filter objects
pizzahut
sys
Administration Schemas
Information
No object selected
Object Info Session
SQL File 1 Retail_Database Retail_Database
Limit to 1000 rows
14 LOAD DATA INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/Retail_Data_Transactions.csv'
15 INTO TABLE Sales_Data_Transactions
16 FIELDS terminated by ','
17 LINES terminated by '\n'
18 IGNORE 1 ROWS;
19
20 SELECT * FROM Sales_Data_Transactions LIMIT 10;
Result Grid Filter Rows: Export: Wrap Cell Content: Fetch rows:
customer_id trans_date tran_amount
CS5295 11-Feb-13 35
CS4768 15-Mar-15 39
CS2122 26-Feb-13 52
CS1217 16-Nov-11 99
CS1850 20-Nov-13 78
CS5539 26-Mar-14 81
CS2724 06-Feb-12 93
CS5902 30-Jan-15 89
CS6040 08-Jan-13 76
CS3802 20-Aug-13 75
Result Grid
Form Editor
Field Types
Transactions 1
Read Only
Context Help Snippets
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Output
Action Output
Time Action Message Duration / Fetch
1 15:21:11 CREATE DATABASE RetailSalesData 1 row(s) affected 0.015 sec
2 15:21:15 USE RetailSalesData 0 row(s) affected 0.000 sec
3 15:21:21 CREATE TABLE Sales_Data_Transactions (customer_id VARCHAR(255), trans_date VARCHAR(255), tran_am... 0 row(s) affected 0.032 sec
4 15:21:41 CREATE TABLE Sales_Data_Response (customer_id VARCHAR(255) PRIMARY KEY, response INT) 0 row(s) affected 0.032 sec
5 15:21:54 LOAD DATA INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/Retail_Data_Transactions.csv' INT... 125000 row(s) affected Records: 125000 Deleted: 0 Skipped: 0 Warnings: 0 1.219 sec
6 15:22:30 SELECT * FROM Sales_Data_Transactions LIMIT 10 10 row(s) returned 0.000 sec / 0.000 sec
Activate Windows
Go to Settings to activate Windows.

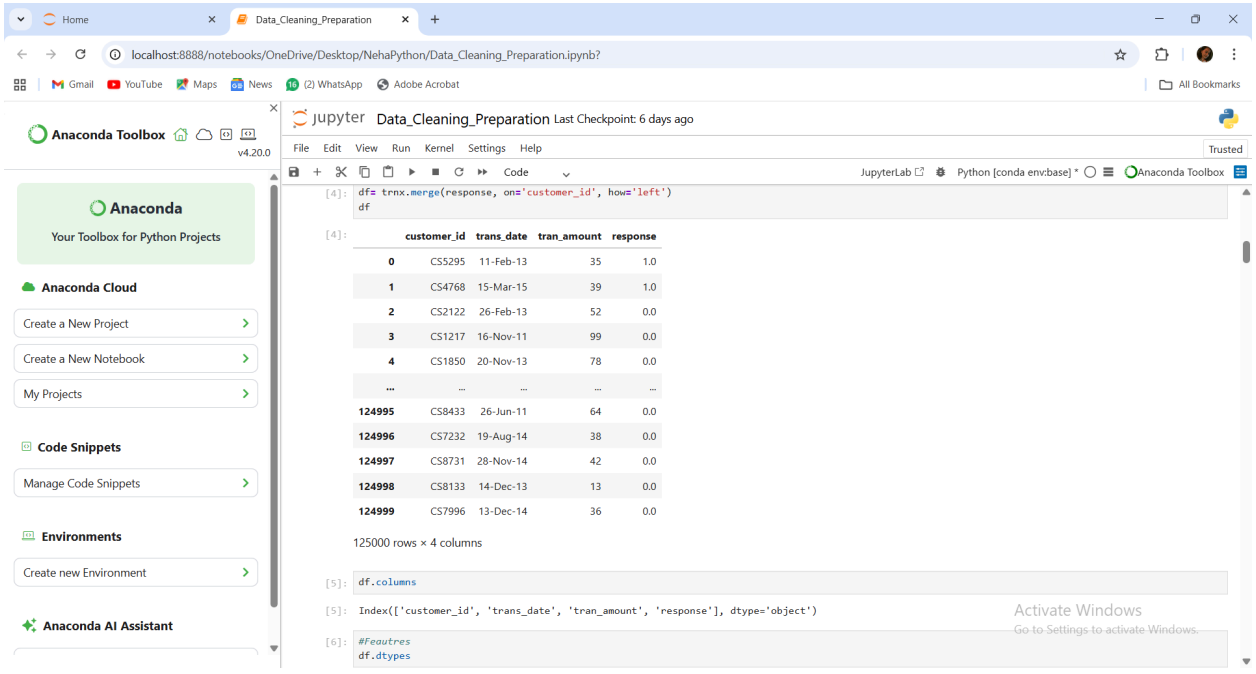

MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
pizzahut
sys
SQL File 1 Retail_Database Retail_Database
Limit to 1000 rows
20 SELECT * FROM Sales_Data_Transactions LIMIT 10;
22 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295';
24 CREATE INDEX idx_id ON Sales_Data_Transactions (CUSTOMER_ID);
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
Fetch rows:
customer_id trans_date tran_amount
CS5295 11-Feb-13 35
CS4768 15-Mar-15 39
CS2122 26-Feb-13 52
CS1217 16-Nov-11 99
CS1850 20-Nov-13 78
CS5539 26-Mar-14 81
CS2724 06-Feb-12 93
CS5902 30-Jan-15 89
CS6040 08-Jan-13 76
CS3802 20-Aug-13 75
Result Grid
Form Editor
Field Types
Administration Schemas
Information
No object selected
Transactions 1
Read Only
Context Help
Snippets
Output
Action Output
Time Action Message Duration / Fetch
1 15:21:11 CREATE DATABASE RetailSalesData 1 row(s) affected 0.015 sec
2 15:21:15 USE RetailSalesData 0 row(s) affected 0.000 sec
3 15:21:21 CREATE TABLE Sales_Data_Transactions (customer_id VARCHAR(255), trans_date VARCHAR(255), tran_am... 0 row(s) affected 0.032 sec
4 15:21:41 CREATE TABLE Sales_Data_Response (customer_id VARCHAR(255) PRIMARY KEY, response INT) 0 row(s) affected 0.032 sec
5 15:21:54 LOAD DATA INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/Retail_Data_Transactions.csv' INT... 125000 row(s) affected Records: 125000 Deleted: 0 Skipped: 0 Warnings: 0 1.219 sec
6 15:22:30 SELECT * FROM Sales_Data_Transactions LIMIT 10 10 row(s) returned 0.000 sec / 0.000 sec
Activate Windows
Go to Settings to activate Windows.
Object Info Session

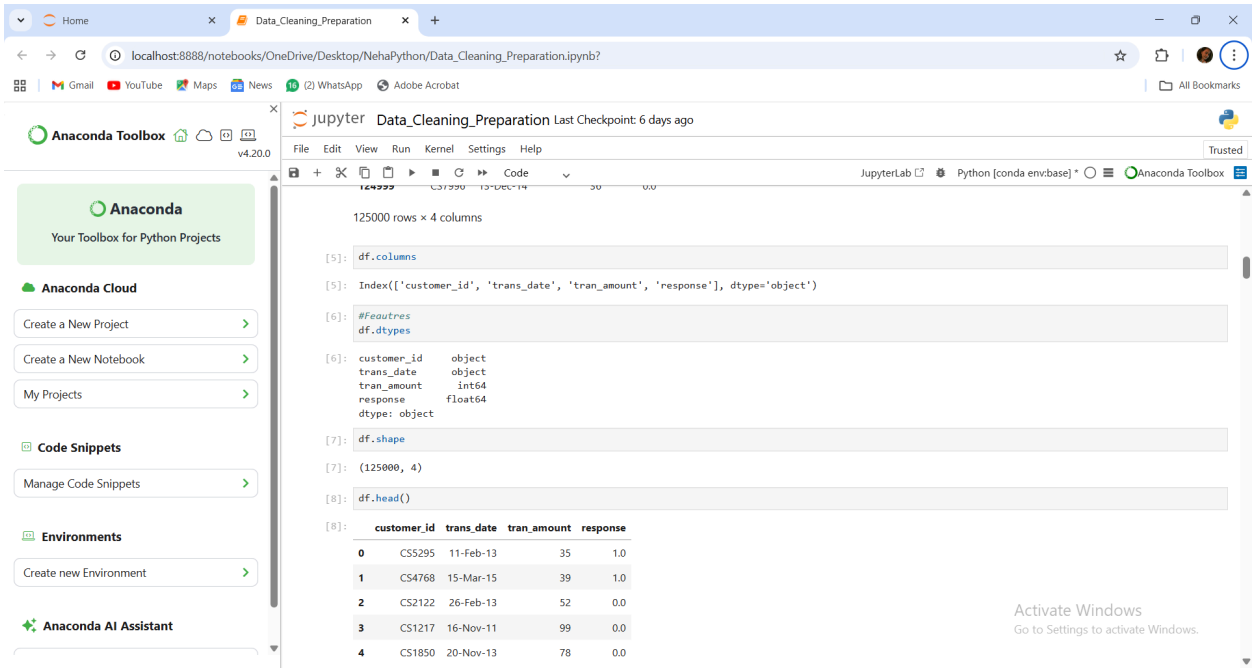

MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
Filter objects
pizzahut
sys
SQL File 1 Retail_Database Retail_Database
Limit to 1000 rows
16 FIELDS terminated by ','
17 LINES terminated by '\n'
18 IGNORE 1 ROWS;
20 SELECT * FROM Sales_Data_Transactions LIMIT 10;
22 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295';
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
id select_type table partitions type possible_keys key key_len ref rows filtered Extra
1 SIMPLE Sales_Data_Transactions NULL ALL NULL NULL NULL NULL 124858 10.00 Using where
Result Grid
Form Editor
Field Types
Administration Schemas
Information
No object selected
Result 2
Read Only
Context Help
Snippets
Output
Action Output
Time Action Message Duration / Fetch
2 15:21:15 USE RetailSalesData 0 row(s) affected 0.000 sec
3 15:21:21 CREATE TABLE Sales_Data_Transactions (customer_id VARCHAR(255), trans_date VARCHAR(255), tran_a... 0 row(s) affected 0.032 sec
4 15:21:41 CREATE TABLE Sales_Data_Response (customer_id VARCHAR(255) PRIMARY KEY, response INT) 0 row(s) affected 0.032 sec
5 15:21:54 LOAD DATA INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/Retail_Data_Transactions.csv' IN... 125000 row(s) affected Records: 125000 Deleted: 0 Skipped: 0 Warnings: 0 1.219 sec
6 15:22:30 SELECT * FROM Sales_Data_Transactions LIMIT 10 10 row(s) returned 0.000 sec / 0.000 sec
7 15:22:55 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295' 1 row(s) returned 0.000 sec / 0.000 sec
Activate Windows
Go to Settings to activate Windows.
Object Info Session

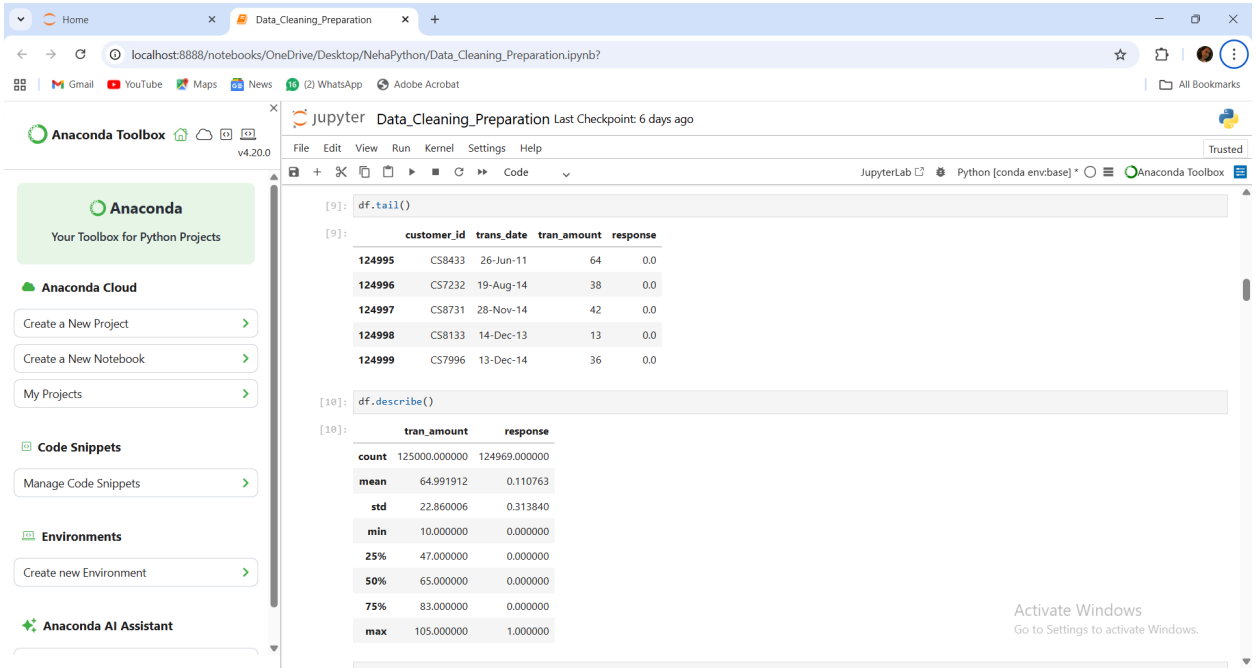

MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator
SCHEMAS
pizzahut
sys
SQL File 1
Retail_Database
Retail_Database
Limit to 1000 rows
19
20 SELECT * FROM Sales_Data_Transactions LIMIT 10;
21
22 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295';
23
24 CREATE INDEX idx_id ON Sales_Data_Transactions (CUSTOMER_ID);
25
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
id select_type table partitions type possible_keys key key_len ref rows filtered Extra
1 SIMPLE Sales_Data_Transactions NULL ALL NULL NULL NULL NULL 124858 10.00 Using where
Administration Schemas
Information
No object selected
Object Info Session
Result 2
Read Only
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
2 15:21:15 USE RetailSalesData 0 row(s) affected 0.000 sec
3 15:21:21 CREATE TABLE Sales_Data_Transactions (customer_id VARCHAR(255), trans_date VARCHAR(255), tran_a... 0 row(s) affected 0.032 sec
4 15:21:41 CREATE TABLE Sales_Data_Response (customer_id VARCHAR(255) PRIMARY KEY, response INT) 0 row(s) affected 0.032 sec
5 15:21:54 LOAD DATA INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/Retail_Data_Transactions.csv' IN... 125000 row(s) affected Records: 125000 Deleted: 0 Skipped: 0 Warnings: 0 1.219 sec
6 15:22:30 SELECT * FROM Sales_Data_Transactions LIMIT 10 10 row(s) returned 0.000 sec / 0.000 sec
7 15:22:55 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295' 1 row(s) returned 0.000 sec / 0.000 sec
Activate Windows
Go to Settings to activate Windows.

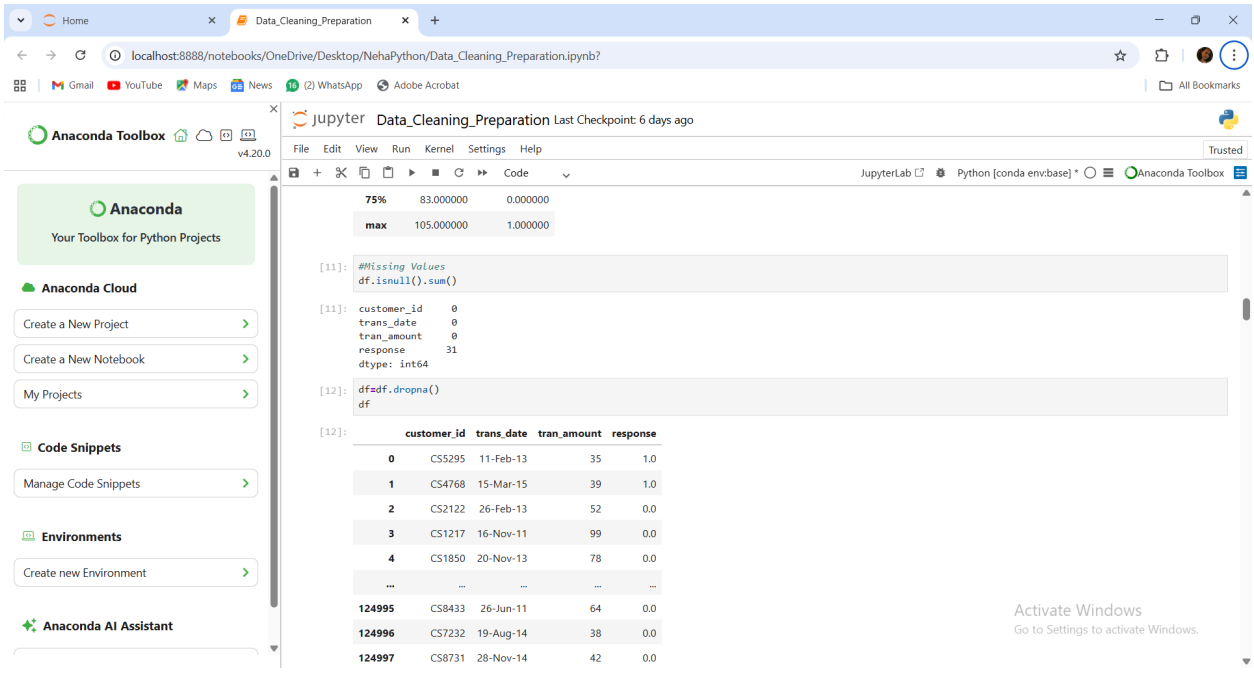

Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
[1]: #Installing Libraries
import pandas as pd
[2]: trnx= pd.read_csv('Retail_Data_Transactions.csv')
trnx
[2]:
customer_id trans_date tran_amount
0 CS5295 11-Feb-13 35
1 CS4768 15-Mar-15 39
2 CS2122 26-Feb-13 52
3 CS1217 16-Nov-11 99
4 CS1850 20-Nov-13 78
... ...
124995 CS8433 26-Jun-11 64
124996 CS7232 19-Aug-14 38
124997 CS8731 28-Nov-14 42
124998 CS8133 14-Dec-13 13
124999 CS7996 13-Dec-14 36
125000 rows × 3 columns
Activate Windows
Go to Settings to activate Windows.

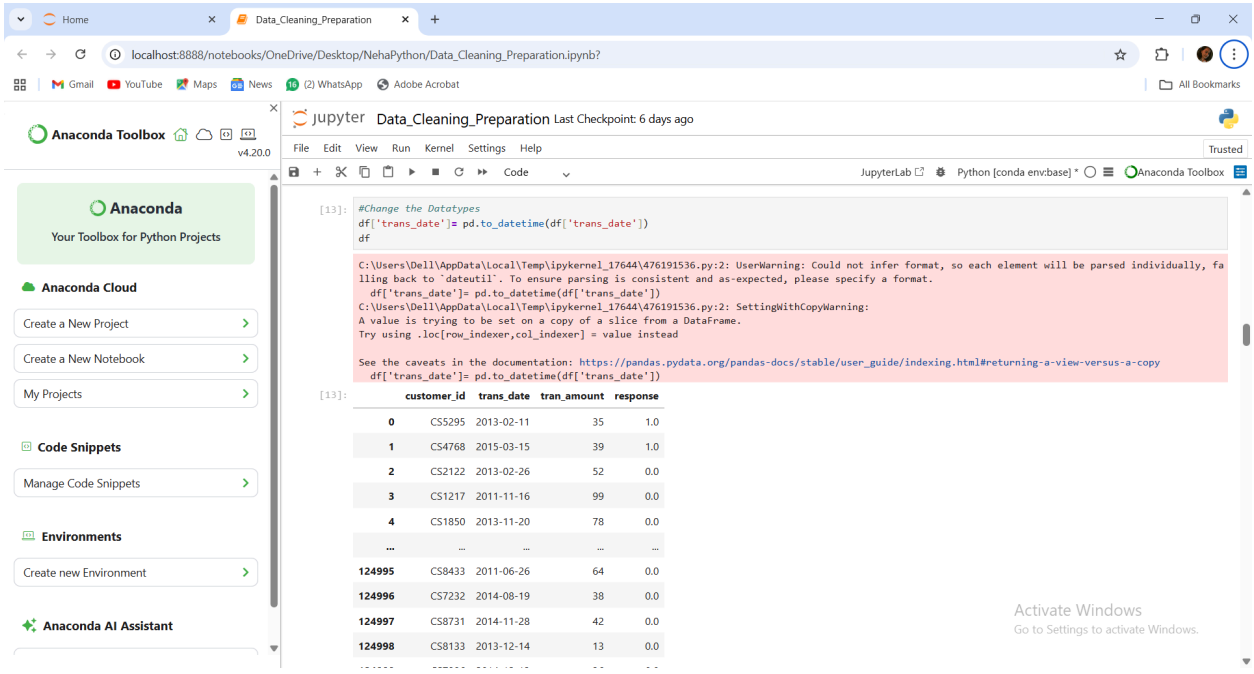

Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat All Bookmarks
Anaconda Toolbox v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
[3]: response= pd.read_csv('Retail_Data_Response.csv')
response
[3]:
customer_id response
0 CS1112 0
1 CS1113 0
2 CS1114 1
3 CS1115 1
4 CS1116 1
... ...
6879 CS8996 0
6880 CS8997 0
6881 CS8998 0
6882 CS8999 0
6883 CS9000 0
6884 rows × 2 columns
[4]: df= trnx.merge(response, on='customer_id', how='left')
df
[4]:
customer_id trans_date tran_amount response
Activate Windows
Go to Settings to activate Windows.

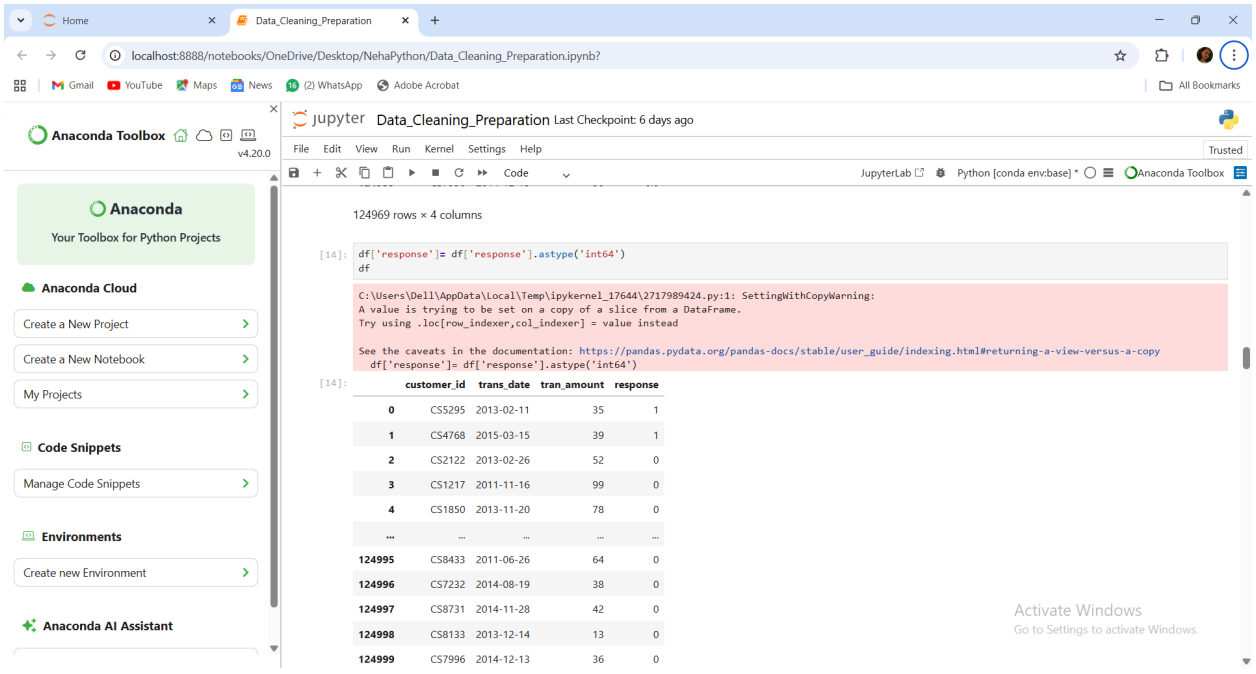

Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat All Bookmarks
Anaconda Toolbox v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
[4]: df= trnx.merge(response, on='customer_id', how='left')
df
[4]:
customer_id trans_date tran_amount response
0 CS5295 11-Feb-13 35 1.0
1 CS4768 15-Mar-15 39 1.0
2 CS2122 26-Feb-13 52 0.0
3 CS1217 16-Nov-11 99 0.0
4 CS1850 20-Nov-13 78 0.0
... ...
124995 CS8433 26-Jun-11 64 0.0
124996 CS7232 19-Aug-14 38 0.0
124997 CS8731 28-Nov-14 42 0.0
124998 CS8133 14-Dec-13 13 0.0
124999 CS7996 13-Dec-14 36 0.0
125000 rows × 4 columns
[5]: df.columns
[5]: Index(['customer_id', 'trans_date', 'tran_amount', 'response'], dtype='object')
[6]: #Feautres
df.dtypes
Activate Windows
Go to Settings to activate Windows.

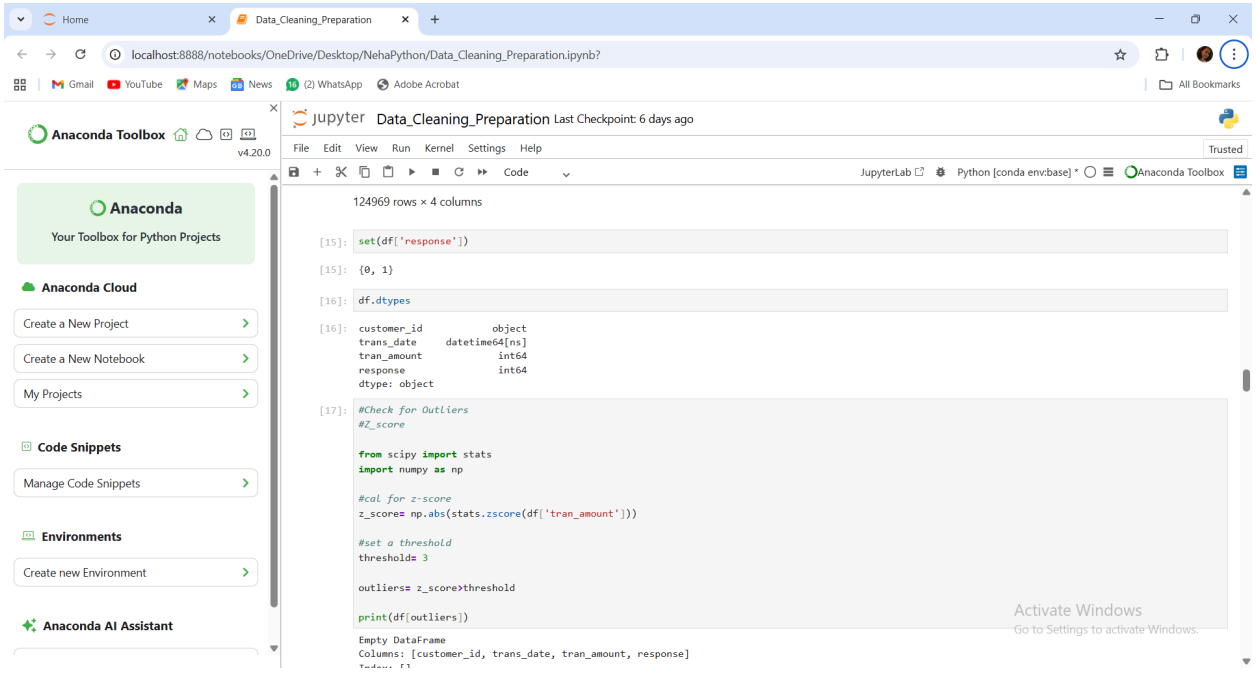

Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter
Data_Cleaning_Preparation
Last Checkpoint: 6 days ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
125000 rows × 4 columns
[5]: df.columns
[5]: Index(['customer_id', 'trans_date', 'tran_amount', 'response'], dtype='object')
[6]: #Feautres
df.dtypes
[6]: customer_id object
trans_date object
tran_amount int64
response float64
dtype: object
[7]: df.shape
[7]: (125000, 4)
[8]: df.head()
[8]:
customer_id trans_date tran_amount response
0 CS5295 11-Feb-13 35 1.0
1 CS4768 15-Mar-15 39 1.0
2 CS2122 26-Feb-13 52 0.0
3 CS1217 16-Nov-11 99 0.0
4 CS1850 20-Nov-13 78 0.0
Activate Windows
Go to Settings to activate Windows.


Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter
Data_Cleaning_Preparation
Last Checkpoint: 6 days ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
[9]: df.tail()
[9]:
customer_id trans_date tran_amount response
124995 CS8433 26-Jun-11 64 0.0
124996 CS7232 19-Aug-14 38 0.0
124997 CS8731 28-Nov-14 42 0.0
124998 CS8133 14-Dec-13 13 0.0
124999 CS7996 13-Dec-14 36 0.0
[10]: df.describe()
[10]:
tran_amount response
count 125000.000000 124969.000000
mean 64.991912 0.110763
std 22.860006 0.313840
min 10.000000 0.000000
25% 47.000000 0.000000
50% 65.000000 0.000000
75% 83.000000 0.000000
max 105.000000 1.000000
Activate Windows
Go to Settings to activate Windows.


Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
75% 83.000000 0.000000
max 105.000000 1.000000
[11]: #Missing Values
df.isnull().sum()
[11]: customer_id 0
trans_date 0
tran_amount 0
response 31
dtype: int64
[12]: df=df.dropna()
df
[12]:
customer_id trans_date tran_amount response
0 CS5295 11-Feb-13 35 1.0
1 CS4768 15-Mar-15 39 1.0
2 CS2122 26-Feb-13 52 0.0
3 CS1217 16-Nov-11 99 0.0
4 CS1850 20-Nov-13 78 0.0
... ...
124995 CS8433 26-Jun-11 64 0.0
124996 CS7232 19-Aug-14 38 0.0
124997 CS8731 28-Nov-14 42 0.0
Activate Windows
Go to Settings to activate Windows.


Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
[13]: #Change the Datatypes
df['trans_date']= pd.to_datetime(df['trans_date'])
df
C:\Users\Dell\AppData\Local\Temp\ipykernel_17644\476191536.py:2: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.
df['trans_date']= pd.to_datetime(df['trans_date'])
C:\Users\Dell\AppData\Local\Temp\ipykernel_17644\476191536.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df['trans_date']= pd.to_datetime(df['trans_date'])
[13]:
customer_id trans_date tran_amount response
0 CS5295 2013-02-11 35 1.0
1 CS4768 2015-03-15 39 1.0
2 CS2122 2013-02-26 52 0.0
3 CS1217 2011-11-16 99 0.0
4 CS1850 2013-11-20 78 0.0
... ...
124995 CS8433 2011-06-26 64 0.0
124996 CS7232 2014-08-19 38 0.0
124997 CS8731 2014-11-28 42 0.0
124998 CS8133 2013-12-14 13 0.0
Activate Windows
Go to Settings to activate Windows.


Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter
Data_Cleaning_Preparation
Last Checkpoint: 6 days ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
124969 rows × 4 columns
[14]:
df['response']= df['response'].astype('int64')
df
C:\Users\Dell\AppData\Local\Temp\ipykernel_17644\2717989424.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df['response']= df['response'].astype('int64')
[14]:
customer_id trans_date tran_amount response
0 CS5295 2013-02-11 35 1
1 CS4768 2015-03-15 39 1
2 CS2122 2013-02-26 52 0
3 CS1217 2011-11-16 99 0
4 CS1850 2013-11-20 78 0
... ...
124995 CS8433 2011-06-26 64 0
124996 CS7232 2014-08-19 38 0
124997 CS8731 2014-11-28 42 0
124998 CS8133 2013-12-14 13 0
124999 CS7996 2014-12-13 36 0
Activate Windows
Go to Settings to activate Windows.


Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter
Data_Cleaning_Preparation
Last Checkpoint: 6 days ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
124969 rows × 4 columns
[15]:
set(df['response'])
[15]:
{0, 1}
[16]:
df.dtypes
[16]:
customer_id object
trans_date datetime64[ns]
tran_amount int64
response int64
dtype: object
[17]:
#Check for Outliers
#Z_score
from scipy import stats
import numpy as np
#cal for z-score
z_score= np.abs(stats.zscore(df['tran_amount']))
#set a threshold
threshold= 3
outliers= z_score>threshold
print(df[outliers])
Empty DataFrame
Columns: [customer_id, trans_date, tran_amount, response]
Index: []
Activate Windows
Go to Settings to activate Windows.

Home | Data_Cleaning_Preparation

localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?

Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat All Bookmarks

Anaconda Toolbox v4.20.0

Anaconda
Your Toolbox for Python Projects

Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects

Code Snippets
Manage Code Snippets

Environments
Create new Environment

Anaconda AI Assistant

jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python [conda env:base] * Anaconda Toolbox

[18]:
```
#Check for Outliers
#Z_score

from scipy import stats
import numpy as np

#cal for z-score
z_score= np.abs(stats.zscore(df['response']))

#set a threshold
threshold= 3

outliers= z_score>threshold

print(df[outliers])
```

```
Empty DataFrame
Columns: [customer_id, trans_date, tran_amount, response]
Index: []
```

[19]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.boxplot(x=df['response'])
plt.show()
```

Activate Windows
Go to Settings to activate Windows.

Home | Data_Cleaning_Preparation

localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?

Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat All Bookmarks

Anaconda Toolbox v4.20.0

Anaconda
Your Toolbox for Python Projects

Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects

Code Snippets
Manage Code Snippets

Environments
Create new Environment

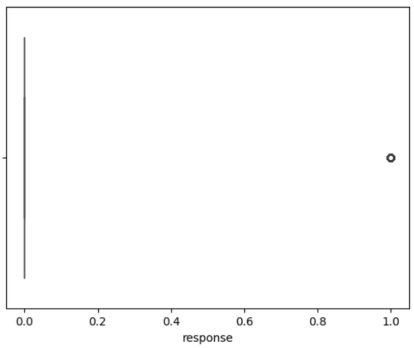
Anaconda AI Assistant

jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python [conda env:base] * Anaconda Toolbox

```
sns.boxplot(x=df['response'])
plt.show()
```



[20]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.boxplot(x=df['tran_amount'])
```

Activate Windows
Go to Settings to activate Windows.

Home | Data_Cleaning_Preparation

localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?

Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat | All Bookmarks

Anaconda Toolbox v4.20.0

Anaconda
Your Toolbox for Python Projects

Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects

Code Snippets
Manage Code Snippets

Environments
Create new Environment

Anaconda AI Assistant

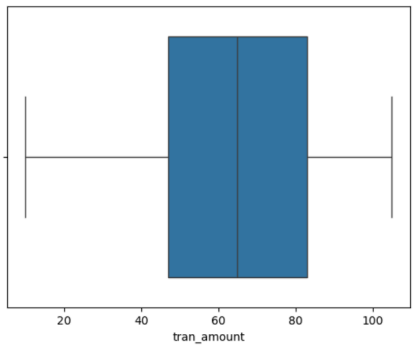
jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago

File Edit View Run Kernel Settings Help

Trusted

Code | JupyterLab | Python [conda env:base] * | Anaconda Toolbox

[20]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.boxplot(x=df['tran_amount'])
plt.show()
```



[21]:
```
#Creating a new columns
```

Home | Data_Cleaning_Preparation

localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?

Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat | All Bookmarks

Anaconda Toolbox v4.20.0

Anaconda
Your Toolbox for Python Projects

Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects

Code Snippets
Manage Code Snippets

Environments
Create new Environment

Anaconda AI Assistant

jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago

File Edit View Run Kernel Settings Help

Trusted

Code | JupyterLab | Python [conda env:base] * | Anaconda Toolbox

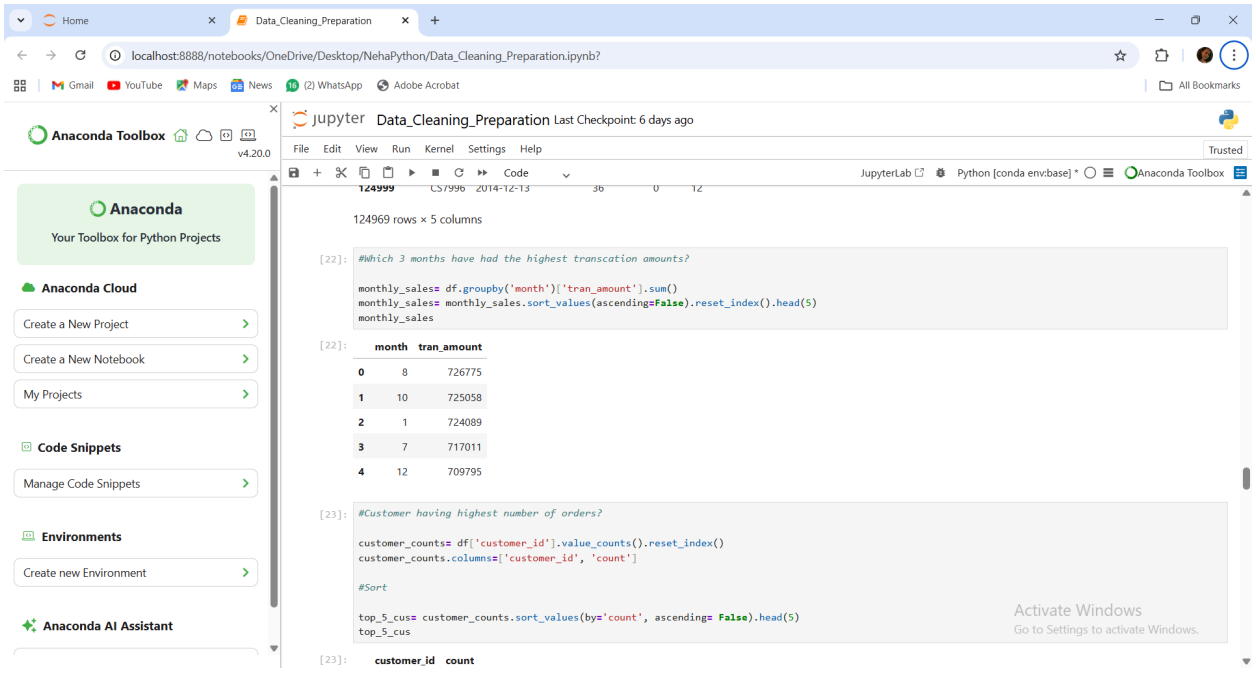
20 40 60 80 100
tran_amount

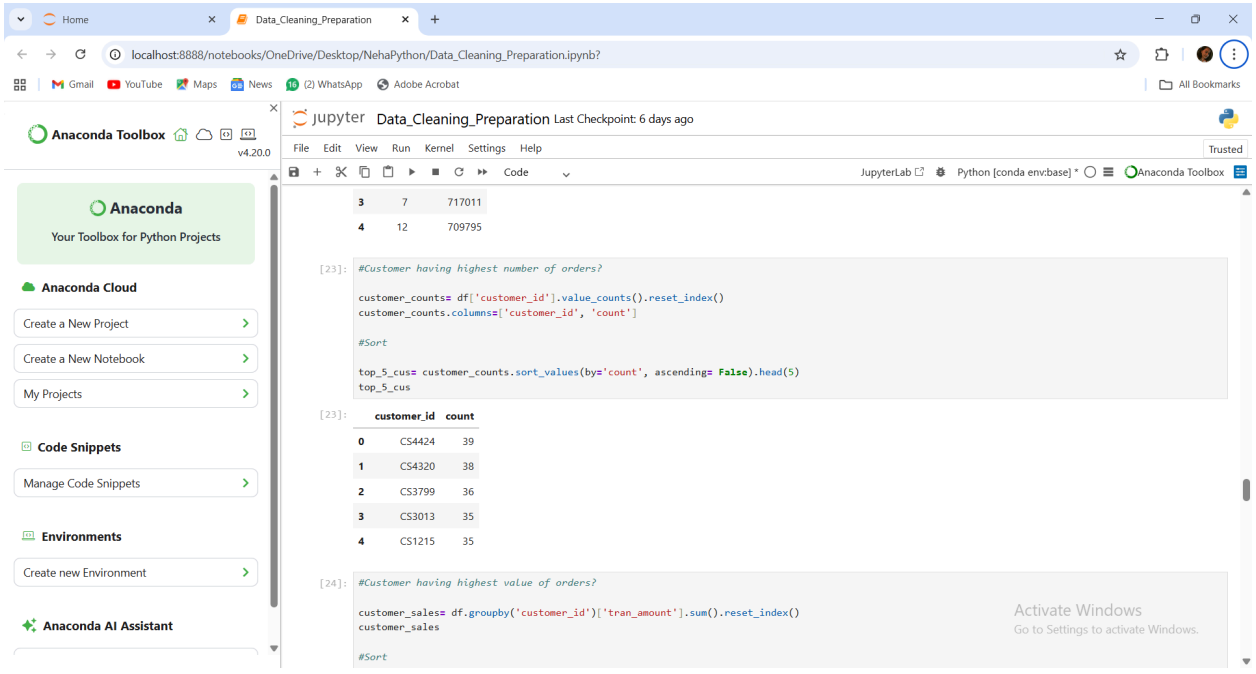
[21]:
```
#Creating a new columns

df['month']= df['trans_date'].dt.month
df
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_17644\48249003.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['month']= df['trans_date'].dt.month
```

[21]:

	customer_id	trans_date	tran_amount	response	month
0	CS5295	2013-02-11	35	1	2
1	CS4768	2015-03-15	39	1	3
2	CS2122	2013-02-26	52	0	2
3	CS1217	2011-11-16	99	0	11
4	CS1850	2013-11-20	78	0	11
...	...	...	...	...	...
124995	CS8433	2011-06-26	64	0	6
124996	CS7232	2014-08-19	38	0	8
124997	CS8731	2014-11-28	42	0	11
124998	CS8133	2013-12-14	13	0	12


Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
124999 CS7996 2014-12-13 36 0 12
124969 rows × 5 columns
[22]:
#Which 3 months have had the highest transcation amounts?
monthly_sales= df.groupby('month')['tran_amount'].sum()
monthly_sales= monthly_sales.sort_values(ascending=False).reset_index().head(5)
monthly_sales
[22]:
month tran_amount
0 8 726775
1 10 725058
2 1 724089
3 7 717011
4 12 709795
[23]:
#Customer having highest number of orders?
customer_counts= df['customer_id'].value_counts().reset_index()
customer_counts.columns=['customer_id', 'count']
#Sort
top_5_cus= customer_counts.sort_values(by='count', ascending= False).head(5)
top_5_cus
Activate Windows
Go to Settings to activate Windows.
[23]:
customer_id count


Home
Data_Cleaning_Preparation
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
3 7 717011
4 12 709795
[23]:
#Customer having highest number of orders?
customer_counts= df['customer_id'].value_counts().reset_index()
customer_counts.columns=['customer_id', 'count']
#Sort
top_5_cus= customer_counts.sort_values(by='count', ascending= False).head(5)
top_5_cus
[23]:
customer_id count
0 CS4424 39
1 CS4320 38
2 CS3799 36
3 CS3013 35
4 CS1215 35
[24]:
#Customer having highest value of orders?
customer_sales= df.groupby('customer_id')['tran_amount'].sum().reset_index()
customer_sales
#Sort
Activate Windows
Go to Settings to activate Windows.

Home | Data_Cleaning_Preparation | New Tab

localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?

Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat All Bookmarks

Anaconda Toolbox v4.20.0

Anaconda
Your Toolbox for Python Projects

Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects

Code Snippets
Manage Code Snippets

Environments
Create new Environment

Anaconda AI Assistant

jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago

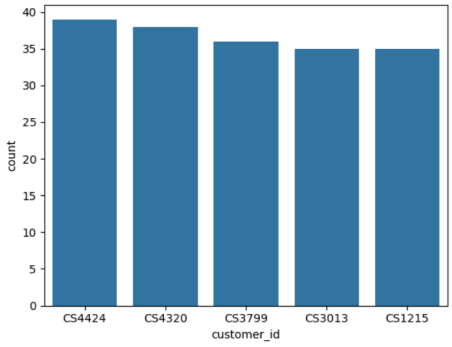
File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python [conda env:base] * Anaconda Toolbox

2687	CS3799	2513

[25]:
```
sns.barplot(x='customer_id',y='count', data=top_5_cus)
```

[25]:
```
<Axes: xlabel='customer_id', ylabel='count'>
```



[26]:
```
sns.barplot(x='customer_id',y='tran_amount', data=top_5_sal)
```

Home | Data_Cleaning_Preparation | New Tab

localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?

Gmail YouTube Maps News (2) WhatsApp Adobe Acrobat All Bookmarks

Anaconda Toolbox v4.20.0

Anaconda
Your Toolbox for Python Projects

Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects

Code Snippets
Manage Code Snippets

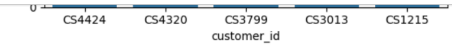
Environments
Create new Environment

Anaconda AI Assistant

jupyter Data_Cleaning_Preparation Last Checkpoint: 6 days ago

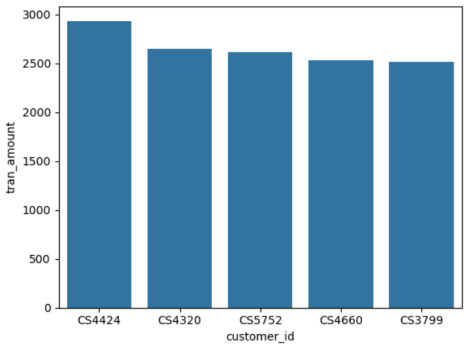
File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python [conda env:base] * Anaconda Toolbox

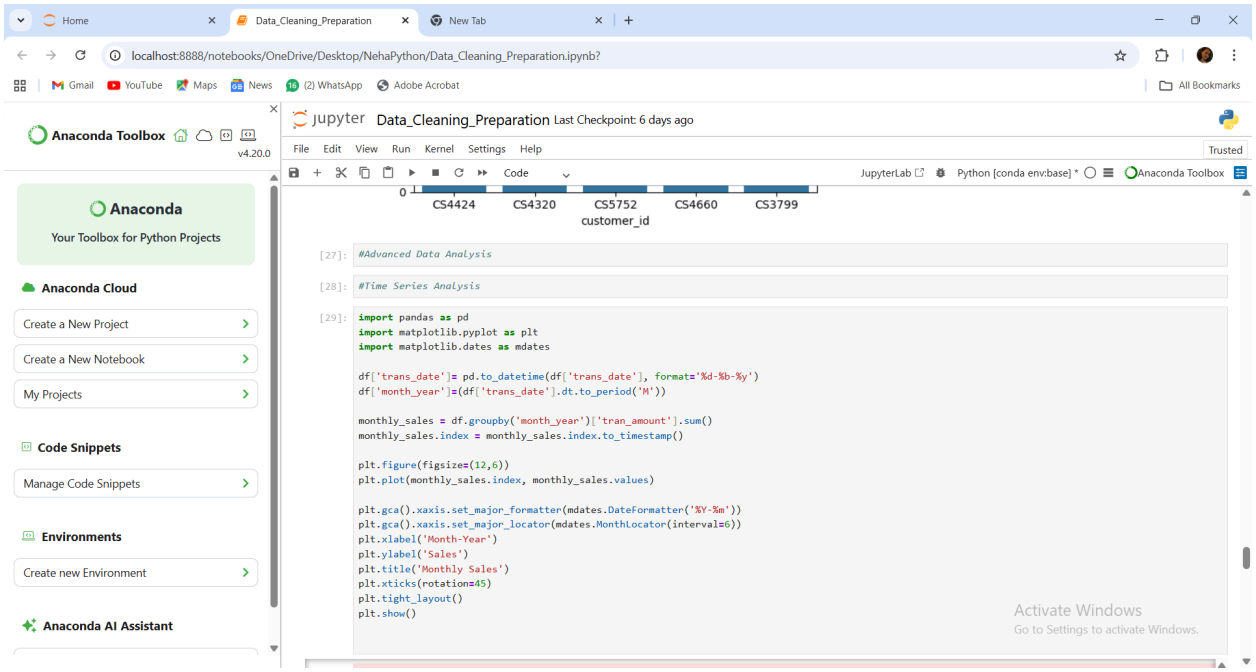


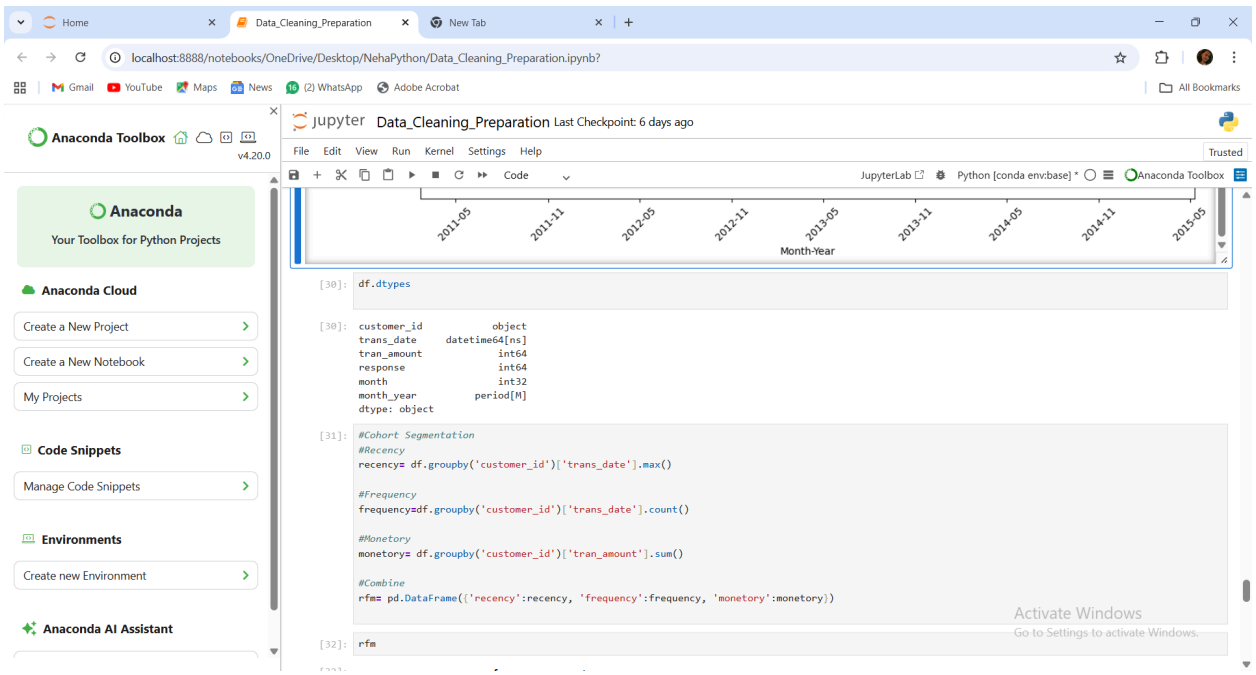
[26]:
```
sns.barplot(x='customer_id',y='tran_amount', data=top_5_sal)
```

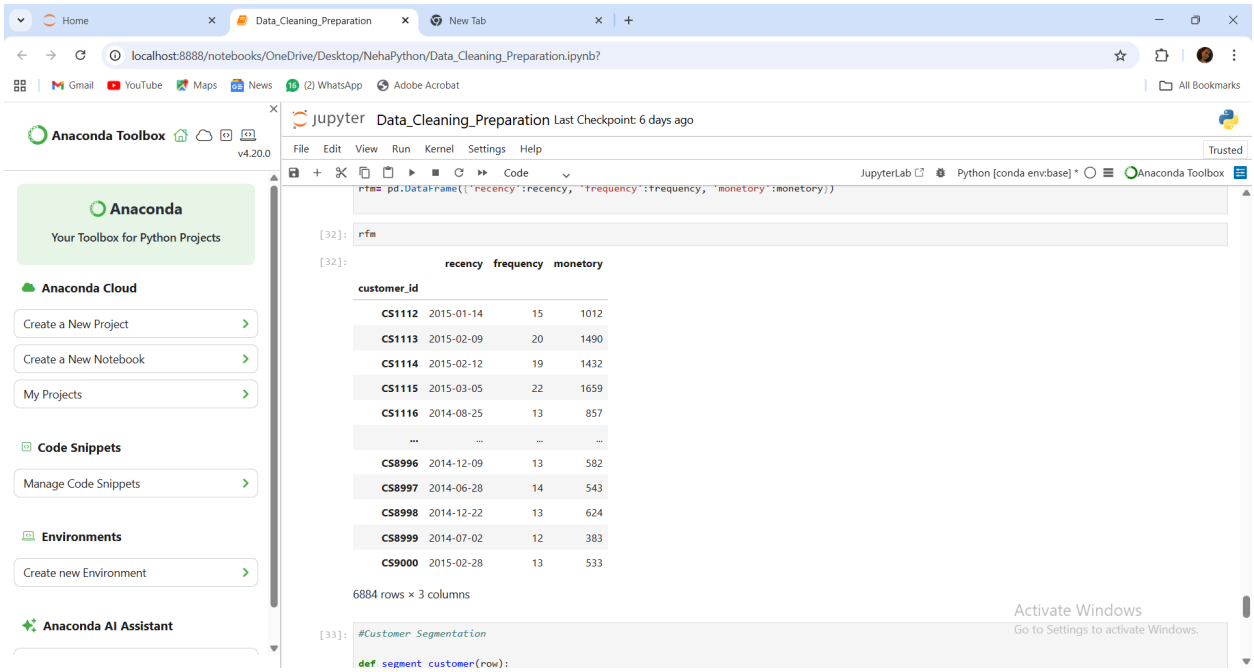
[26]:
```
<Axes: xlabel='customer_id', ylabel='tran_amount'>
```

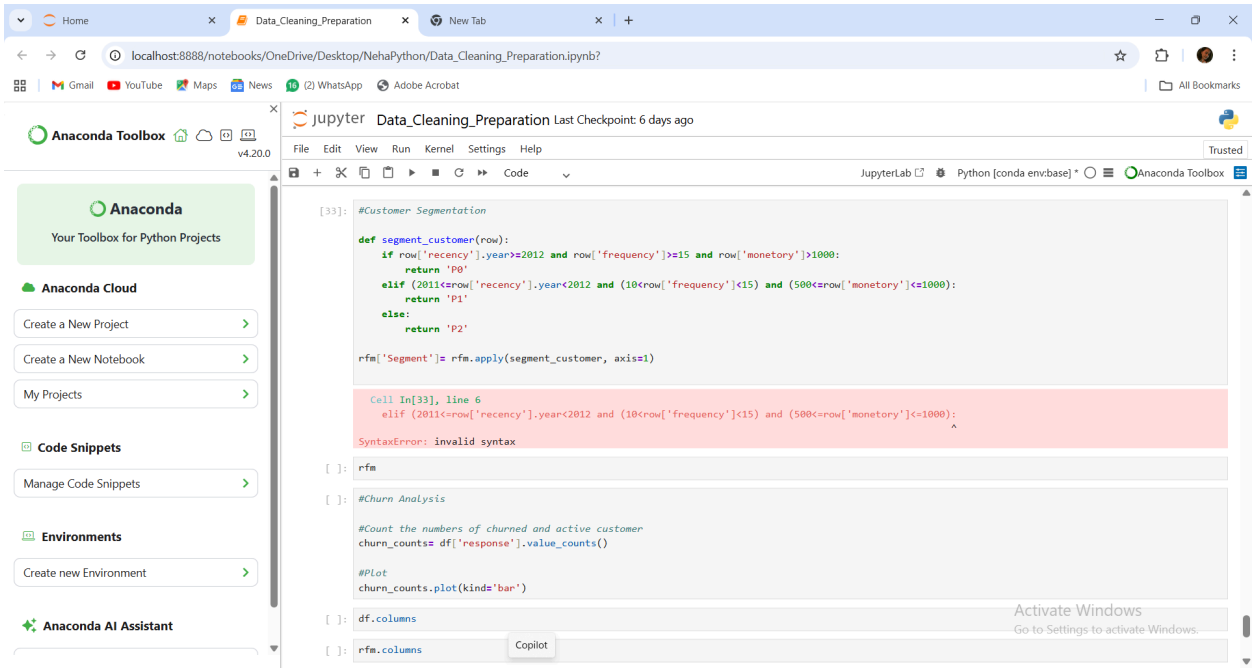


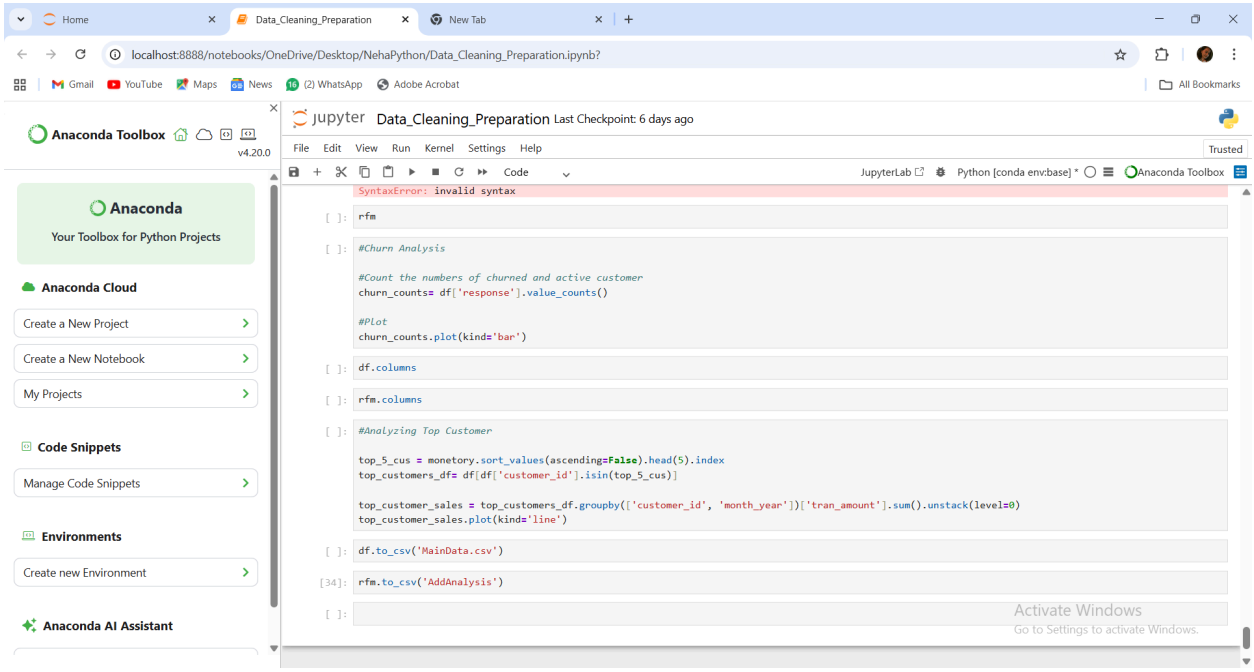
[27]:
```
#Advanced Data Analysis
```


Home
Data_Cleaning_Preparation
New Tab
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter
Data_Cleaning_Preparation
Last Checkpoint: 6 days ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
0
CS4424
CS4320
CS5752
CS4660
CS3799
customer_id
[27]:
#Advanced Data Analysis
[28]:
#Time Series Analysis
[29]:
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
df['trans_date']= pd.to_datetime(df['trans_date'], format='%d-%b-%y')
df['month_year']=(df['trans_date'].dt.to_period('M'))
monthly_sales = df.groupby('month_year')['tran_amount'].sum()
monthly_sales.index = monthly_sales.index.to_timestamp()
plt.figure(figsize=(12,6))
plt.plot(monthly_sales.index, monthly_sales.values)
plt.gca().xaxis.set_major_formatter(mdates.DateFormatter('%Y-%m'))
plt.gca().xaxis.set_major_locator(mdates.MonthLocator(interval=6))
plt.xlabel('Month-Year')
plt.ylabel('Sales')
plt.title('Monthly Sales')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
Activate Windows
Go to Settings to activate Windows.

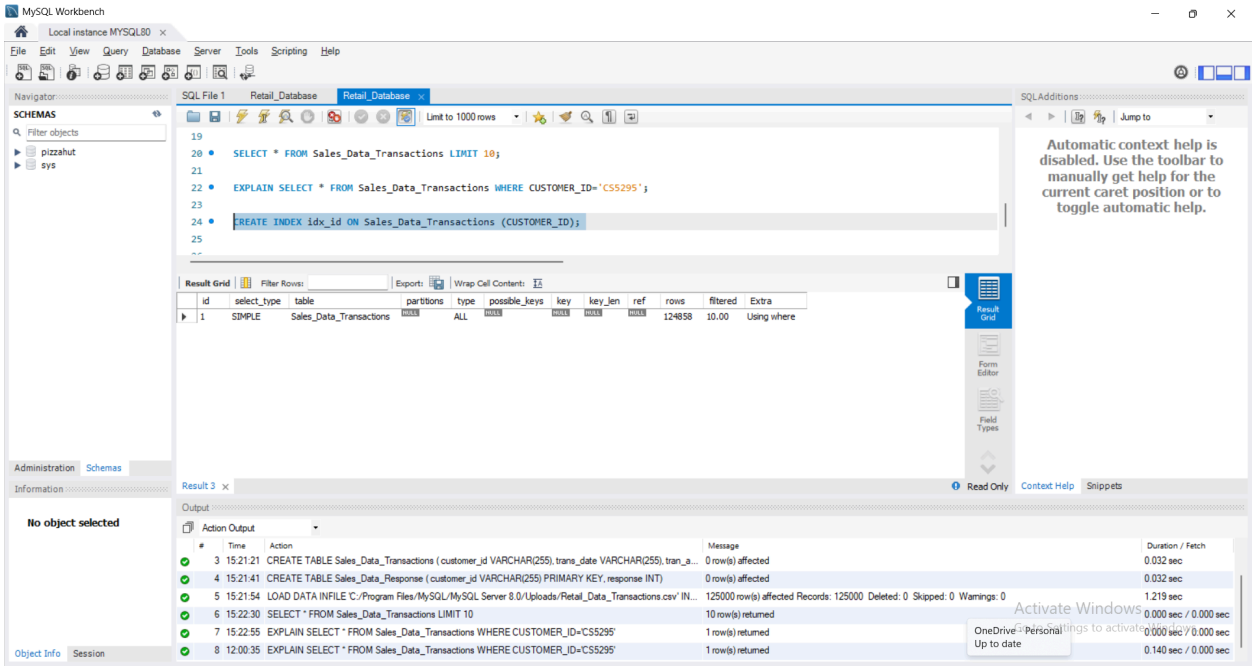

Home
Data_Cleaning_Preparation
New Tab
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter
Data_Cleaning_Preparation
Last Checkpoint: 6 days ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
2011-05
2011-11
2012-05
2012-11
2013-05
2013-11
2014-05
2014-11
2015-05
Month-Year
[30]:
df.dtypes
[30]:
customer_id object
trans_date datetime64[ns]
tran_amount int64
response int64
month int32
month_year period[M]
dtype: object
[31]:
#Cohort Segmentation
#Recency
recency= df.groupby('customer_id')['trans_date'].max()
#Frequency
frequency=df.groupby('customer_id')['trans_date'].count()
#Monetory
monetory= df.groupby('customer_id')['tran_amount'].sum()
#Combine
rfm= pd.DataFrame({'recency':recency, 'frequency':frequency, 'monetory':monetory})
[32]:
rfm
Activate Windows
Go to Settings to activate Windows.

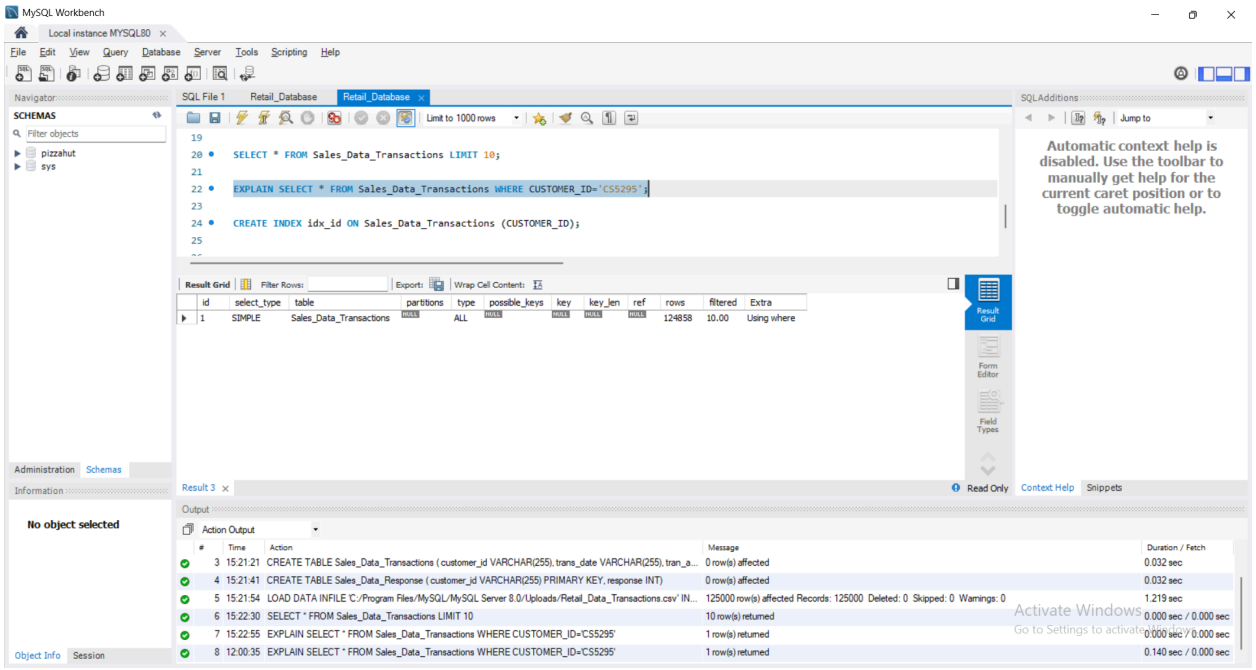

Home
Data_Cleaning_Preparation
New Tab
localhost:8888/notebooks/OneDrive/Desktop/NehaPython/Data_Cleaning_Preparation.ipynb?
Gmail
YouTube
Maps
News
(2) WhatsApp
Adobe Acrobat
All Bookmarks
Anaconda Toolbox
v4.20.0
Anaconda
Your Toolbox for Python Projects
Anaconda Cloud
Create a New Project
Create a New Notebook
My Projects
Code Snippets
Manage Code Snippets
Environments
Create new Environment
Anaconda AI Assistant
jupyter
Data_Cleaning_Preparation
Last Checkpoint: 6 days ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python [conda env:base] *
Anaconda Toolbox
rfm= pd.DataFrame({"recency":recency, "frequency":frequency, "monetory":monetory})
[32]:
rfm
[32]:
recency
frequency
monetory
customer_id
CS1112 2015-01-14 15 1012
CS1113 2015-02-09 20 1490
CS1114 2015-02-12 19 1432
CS1115 2015-03-05 22 1659
CS1116 2014-08-25 13 857
... ...
CS8996 2014-12-09 13 582
CS8997 2014-06-28 14 543
CS8998 2014-12-22 13 624
CS8999 2014-07-02 12 383
CS9000 2015-02-28 13 533
6884 rows × 3 columns
Activate Windows
Go to Settings to activate Windows.
[33]:
#Customer Segmentation
def segment customer(row):

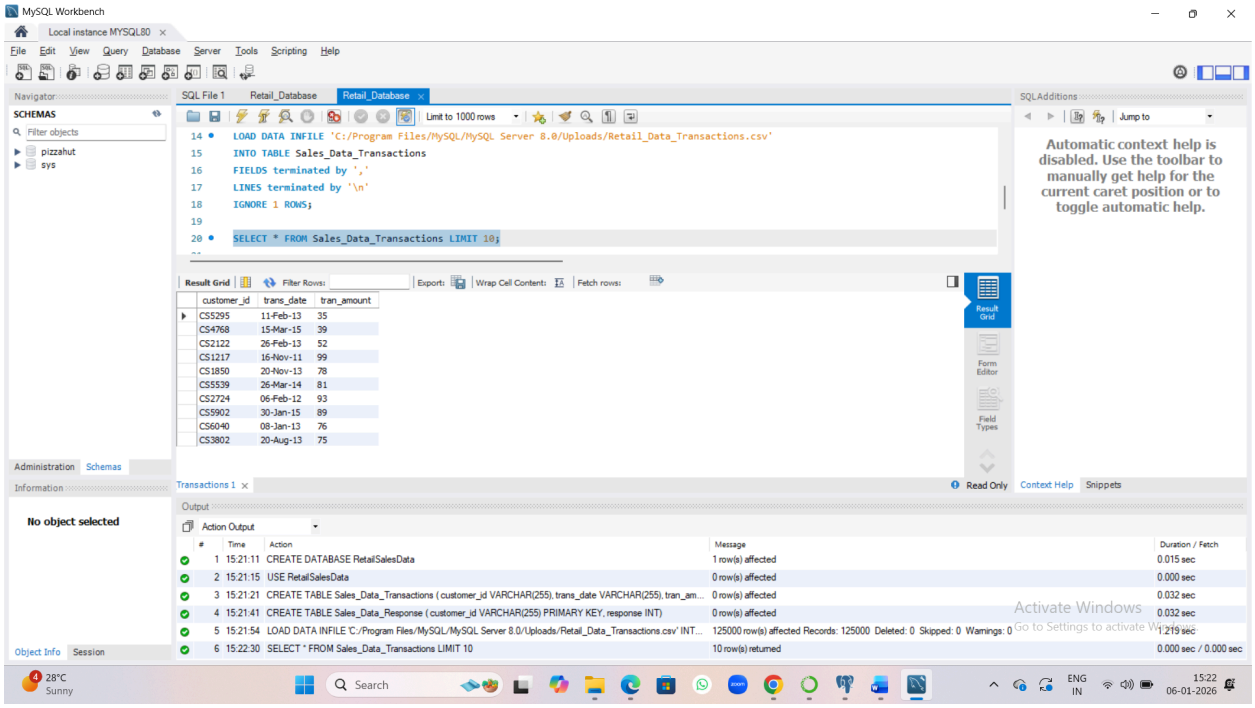




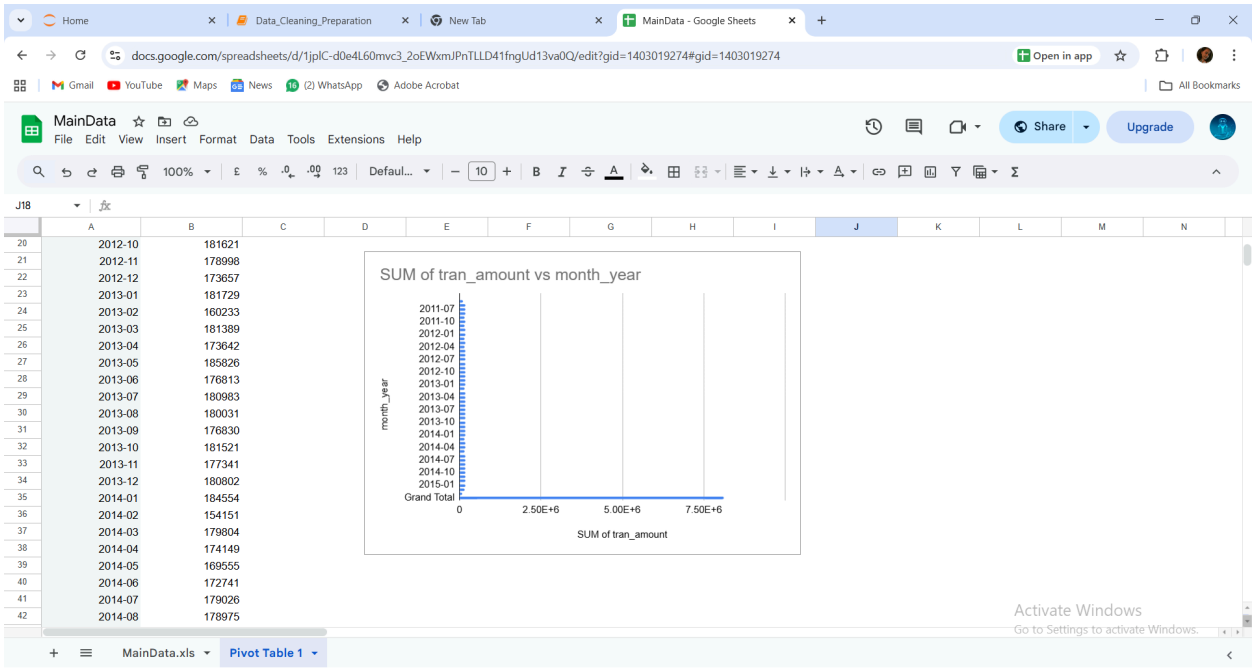
- MySQL query results


MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator:
SCHEMAS
Filter objects
pizzahut
sys
SQL File 1 Retail_Database Retail_Database
Limit to 1000 rows
19
20 SELECT * FROM Sales_Data_Transactions LIMIT 10;
21
22 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295';
23
24 CREATE INDEX idx_id ON Sales_Data_Transactions (CUSTOMER_ID);
25
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Result Grid Filter Rows: Export: Wrap Cell Content:
id select_type table partitions type possible_keys key key_len ref rows filtered Extra
1 SIMPLE Sales_Data_Transactions NULL ALL NULL NULL NULL NULL 124858 10.00 Using where
Result Grid
Form Editor
Field Types
Administration Schemas
Information
No object selected
Result 3
Read Only
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
3 15:21:21 CREATE TABLE Sales_Data_Transactions (customer_id VARCHAR(255), trans_date VARCHAR(255), tran_a... 0 row(s) affected 0.032 sec
4 15:21:41 CREATE TABLE Sales_Data_Response (customer_id VARCHAR(255) PRIMARY KEY, response INT) 0 row(s) affected 0.032 sec
5 15:21:54 LOAD DATA INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/Retail_Data_Transactions.csv' IN... 125000 row(s) affected Records: 125000 Deleted: 0 Skipped: 0 Warnings: 0 1.219 sec
6 15:22:30 SELECT * FROM Sales_Data_Transactions LIMIT 10 10 row(s) returned 0.000 sec / 0.000 sec
7 15:22:55 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295' 1 row(s) returned 0.000 sec / 0.000 sec
8 12:00:35 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295' 1 row(s) returned 0.140 sec / 0.000 sec
Object Info Session
OneDrive - Personal
Up to date
Activate Windows
Go to Settings to activate Windows.

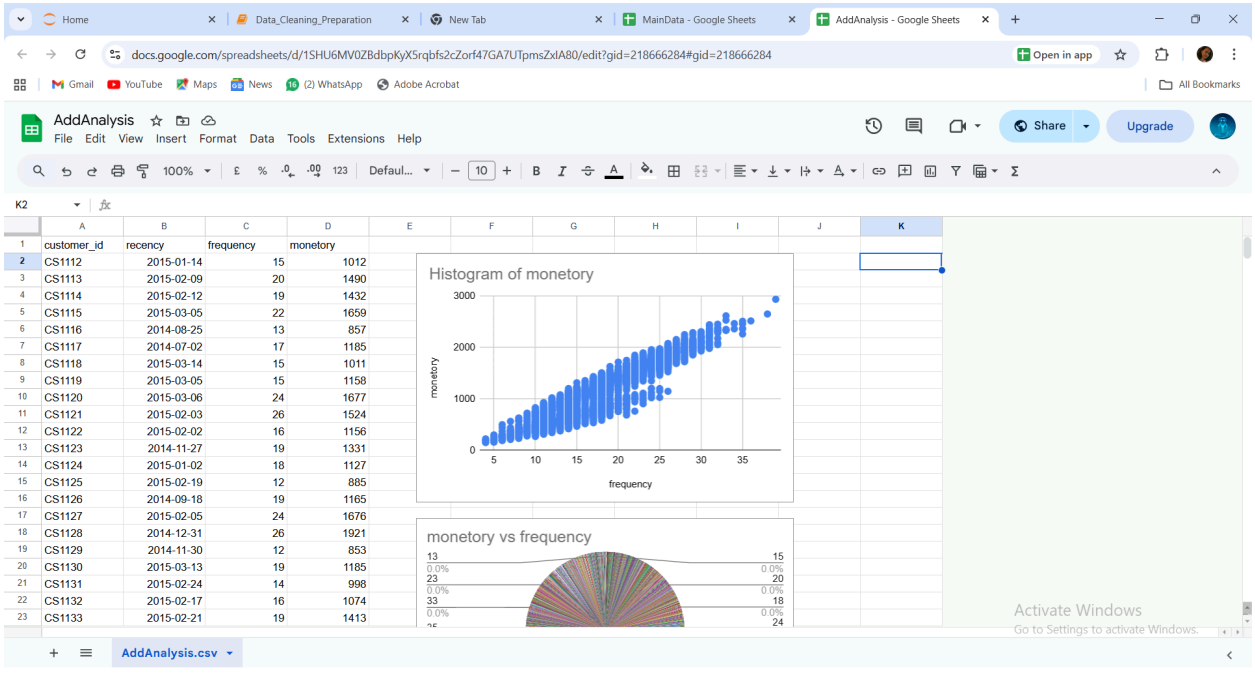

MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator:
SCHEMAS
Filter objects
pizzahut
sys
SQL File 1 Retail_Database Retail_Database
Limit to 1000 rows
19
20 SELECT * FROM Sales_Data_Transactions LIMIT 10;
21
22 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295';
23
24 CREATE INDEX idx_id ON Sales_Data_Transactions (CUSTOMER_ID);
25
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Result Grid Filter Rows: Export: Wrap Cell Content:
id select_type table partitions type possible_keys key key_len ref rows filtered Extra
1 SIMPLE Sales_Data_Transactions NULL ALL NULL NULL NULL NULL 124858 10.00 Using where
Result Grid
Form Editor
Field Types
Administration Schemas
Information
No object selected
Result 3
Read Only
Context Help Snippets
Output
Action Output
Time Action Message Duration / Fetch
3 15:21:21 CREATE TABLE Sales_Data_Transactions (customer_id VARCHAR(255), trans_date VARCHAR(255), tran_a... 0 row(s) affected 0.032 sec
4 15:21:41 CREATE TABLE Sales_Data_Response (customer_id VARCHAR(255) PRIMARY KEY, response INT) 0 row(s) affected 0.032 sec
5 15:21:54 LOAD DATA INFILE 'C:/Program Files/MySQL/MySQL Server 8.0/Uploads/Retail_Data_Transactions.csv' IN... 125000 row(s) affected Records: 125000 Deleted: 0 Skipped: 0 Warnings: 0 1.219 sec
6 15:22:30 SELECT * FROM Sales_Data_Transactions LIMIT 10 10 row(s) returned 0.000 sec / 0.000 sec
7 15:22:55 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295' 1 row(s) returned 0.000 sec / 0.000 sec
8 12:00:35 EXPLAIN SELECT * FROM Sales_Data_Transactions WHERE CUSTOMER_ID='CS5295' 1 row(s) returned 0.140 sec / 0.000 sec
Object Info Session
Activate Windows
Go to Settings to activate Windows.



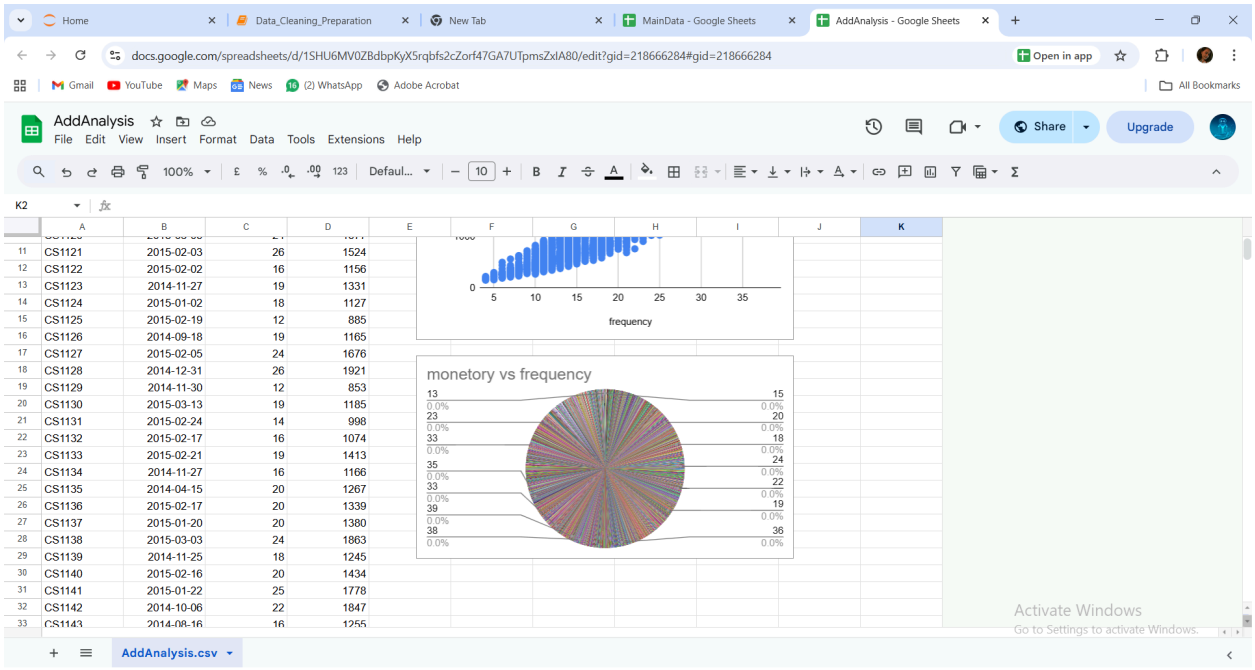
- Google Sheets pivot tables and charts



	A	B
20	2012-10	181621
21	2012-11	178998
22	2012-12	173657
23	2013-01	181729
24	2013-02	160233
25	2013-03	181389
26	2013-04	173642
27	2013-05	185826
28	2013-06	176813
29	2013-07	180983
30	2013-08	180031
31	2013-09	176830
32	2013-10	181521
33	2013-11	177341
34	2013-12	180802
35	2014-01	184554
36	2014-02	154151
37	2014-03	179804
38	2014-04	174149
39	2014-05	169555
40	2014-06	172741
41	2014-07	179026
42	2014-08	178975



	A	B	C	D
1	customer_id	recency	frequency	monetory
2	CS1112	2015-01-14	15	1012
3	CS1113	2015-02-09	20	1490
4	CS1114	2015-02-12	19	1432
5	CS1115	2015-03-05	22	1659
6	CS1116	2014-08-25	13	857
7	CS1117	2014-07-02	17	1185
8	CS1118	2015-03-14	15	1011
9	CS1119	2015-03-05	15	1158
10	CS1120	2015-03-06	24	1677
11	CS1121	2015-02-03	26	1524
12	CS1122	2015-02-02	16	1156
13	CS1123	2014-11-27	19	1331
14	CS1124	2015-01-02	18	1127
15	CS1125	2015-02-19	12	885
16	CS1126	2014-09-18	19	1165
17	CS1127	2015-02-05	24	1676
18	CS1128	2014-12-31	26	1921
19	CS1129	2014-11-30	12	853
20	CS1130	2015-03-13	19	1185
21	CS1131	2015-02-24	14	998
22	CS1132	2015-02-17	16	1074
23	CS1133	2015-02-21	19	1413



Summary

This project focuses on analyzing inventory and stock movement data to identify slow-moving items, overstock risks, and stock-out risks. It demonstrates inventory KPIs using SQL, Python, and Google Sheets pivot reports.