Team Details

Presented By:

Surapuraju Jagadeeswar Raju (2203031050673)

Madhuri Ashok Reddy (2203031050328)

Rohit Lakhamanbhai Vala (220303105782)

Vivek Vinod Prasad (2203031050804)

Guide:
Assistant Professor Akshara Prachi

Data-Driven Customer Segmentation for E-Commerce and Retail Optimization

Project Abstract:

- Analyze customer data from e-commerce and retail businesses.
- Use K-Means Clustering to segment customers.
- Identify customer groups based on behavior, preferences, and demographics.
- Help businesses personalize marketing strategies and improve customer satisfaction

Why Data-Driven Segmentation?

1 Outdated Methods

Traditional segmentation methods are inefficient and outdated.

2 Targeted Marketing

Businesses struggle to tailor their marketing campaigns effectively.

3 Business Impact

Expected business impact: higher conversion rates, better engagement, and increased revenue.

Project Execution Flow

1

Data Collection

Gathering data from e-commerce platforms, CRM systems, and customer interactions.

2

Data Preprocessing

Cleaning data, handling missing values, and normalizing features.

3

Feature Engineering

Creating relevant features like RFM scores, customer engagement scores, and purchasing patterns.

4

Segmentation Techniques

Applying clustering algorithms to categorize customers.

5

Model Evaluation

Validating cluster quality and business relevance.

6

Business Implementation

Deploying segmentation insights for marketing and sales strategies.

7

Performance Monitoring

Updating models based on changing customer behavior.

Data Collection

Transactional Data

Purchase frequency, order value, refunds.

Demographic Data

Age, location, income group.

Behavioral Data

Website interactions, email clicks, ad response.

Engagement Data

Participation in loyalty programs, social media activity.

Data Preprocessing

1

Missing Values

Remove or impute missing values using mean, median, or predictive models.

2

Duplicates

Identify and eliminate duplicate records to avoid redundancy.

3

Outliers

Use IQR method or Z-score analysis to handle anomalies.

4

Standardization & Normalization

Ensure uniformity in data formats, currency conversion, and categorical encoding.

Feature Engineering

RFM Analysis

Recency: How recently a customer made a purchase. Frequency: How often they buy. Monetary: How much they spend.

Customer Lifetime Value (CLV)

Predicting revenue contribution over the customer's lifetime.

Engagement Scores

Analyzing email interactions, click rates, and social media engagement.

Segmentation Techniques

RFM Analysis

Categorizing customers into high-value, at-risk, and lost customers.

K-Means Clustering

Assigning customers to groups based on purchasing behavior.

Hierarchical Clustering

Creating nested customer segments for deeper insights.

DBSCAN

Detecting high-density clusters and noise.

Machine Learning

Using AI to dynamically segment customers based on new data.

AI-Driven Segmentation: Implementing the Future of Marketing

Personalized Customer Engagement

AI-powered segmentation allows businesses to identify customer groups with unique needs, delivering personalized experiences that drive engagement and conversions.

Data-Driven Insights

By analyzing customer data, AI models uncover hidden patterns and insights that inform marketing strategies for better targeting and impactful campaigns.

Boosting Customer Retention

Understanding customer behavior through AI segmentation empowers businesses to create loyalty programs and targeted campaigns that retain valuable customers.

Continuous Monitoring and Optimization

1. Segmentation models require constant monitoring and optimization to ensure accuracy and effectiveness.
2. Continuous learning allows models to adapt to seasonal changes and emerging market trends.
3. Updating segmentation strategies based on shifting customer behaviors is crucial for campaign success.

Real-Time Segmentation: Adapting to Changing Behavior

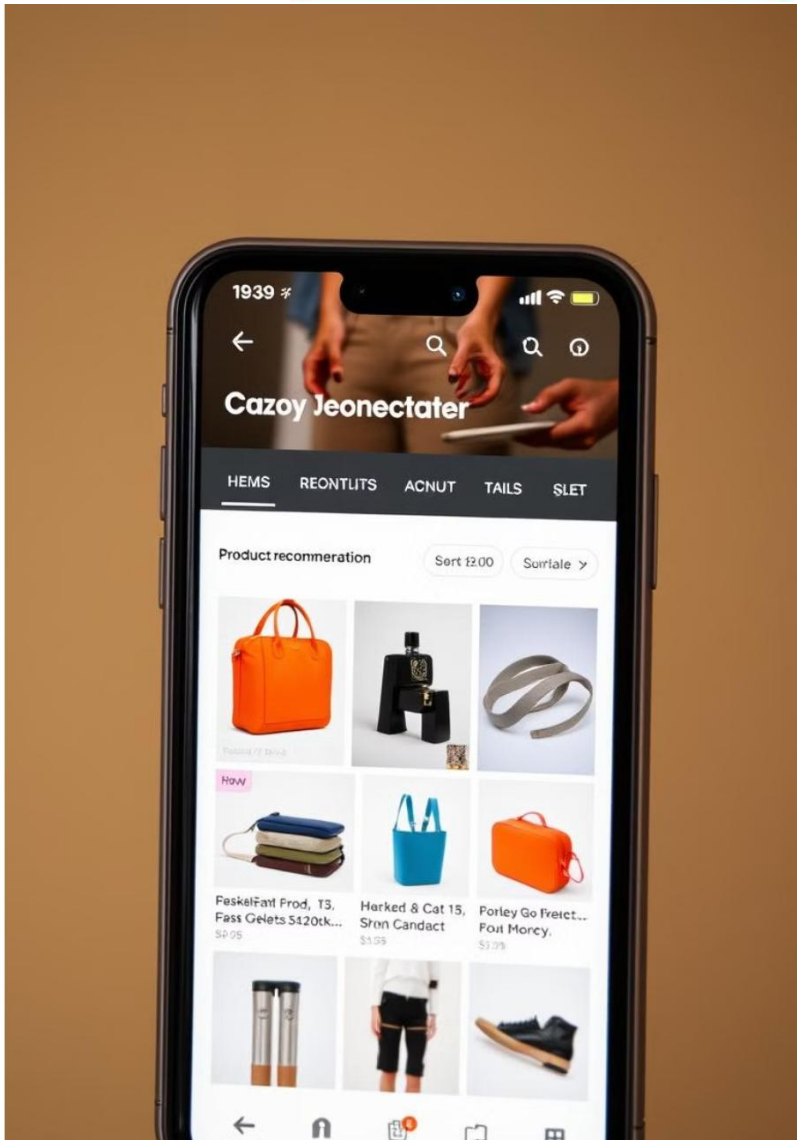
- Real-time segmentation updates based on new customer behaviors.
- Continuous learning models adapt to recent transactions and interactions.
- Dynamic segmentation vs. static segmentation.
- Implementation challenges and solutions.

Personalization through AI-Powered Recommendations

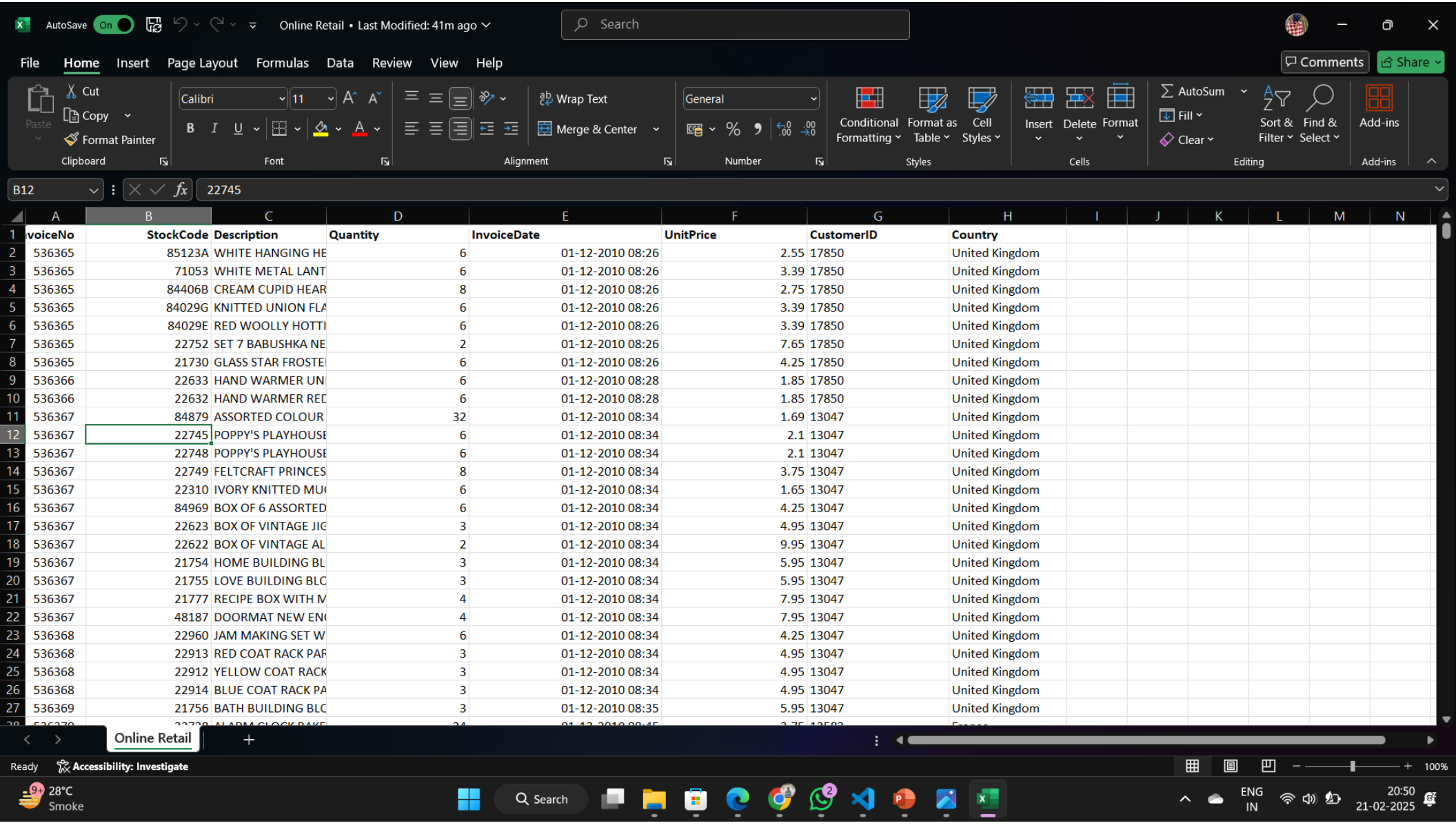
Recommendation engines provide personalized product suggestions, enhancing user experience.

Examples include Netflix and Amazon, which use AI to personalize content and product recommendations.

Personalized suggestions significantly impact customer retention and conversion rates, boosting business growth.



Project Execution Screenshots



	A	B	C	D	E	F	G	H
1	voiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
2	536365	85123A	WHITE HANGING HE	6	01-12-2010 08:26	2.55	17850	United Kingdom
3	536365	71053	WHITE METAL LANT	6	01-12-2010 08:26	3.39	17850	United Kingdom
4	536365	84406B	CREAM CUPID HEAR	8	01-12-2010 08:26	2.75	17850	United Kingdom
5	536365	84029G	KNITTED UNION FLA	6	01-12-2010 08:26	3.39	17850	United Kingdom
6	536365	84029E	RED WOOLLY HOTTI	6	01-12-2010 08:26	3.39	17850	United Kingdom
7	536365	22752	SET 7 BABUSHKA NE	2	01-12-2010 08:26	7.65	17850	United Kingdom
8	536365	21730	GLASS STAR FROSTE	6	01-12-2010 08:26	4.25	17850	United Kingdom
9	536366	22633	HAND WARMER UN	6	01-12-2010 08:28	1.85	17850	United Kingdom
10	536366	22632	HAND WARMER RED	6	01-12-2010 08:28	1.85	17850	United Kingdom
11	536367	84879	ASSORTED COLOUR	32	01-12-2010 08:34	1.69	13047	United Kingdom
12	536367	22745	POPPY'S PLAYHOUSE	6	01-12-2010 08:34	2.1	13047	United Kingdom
13	536367	22748	POPPY'S PLAYHOUSE	6	01-12-2010 08:34	2.1	13047	United Kingdom
14	536367	22749	FELTCRAFT PRINCES	8	01-12-2010 08:34	3.75	13047	United Kingdom
15	536367	22310	IVORY KNITTED MU	6	01-12-2010 08:34	1.65	13047	United Kingdom
16	536367	84969	BOX OF 6 ASSORTED	6	01-12-2010 08:34	4.25	13047	United Kingdom
17	536367	22623	BOX OF VINTAGE JIG	3	01-12-2010 08:34	4.95	13047	United Kingdom
18	536367	22622	BOX OF VINTAGE AL	2	01-12-2010 08:34	9.95	13047	United Kingdom
19	536367	21754	HOME BUILDING BL	3	01-12-2010 08:34	5.95	13047	United Kingdom
20	536367	21755	LOVE BUILDING BLO	3	01-12-2010 08:34	5.95	13047	United Kingdom
21	536367	21777	RECIPE BOX WITH M	4	01-12-2010 08:34	7.95	13047	United Kingdom
22	536367	48187	DOORMAT NEW EN	4	01-12-2010 08:34	7.95	13047	United Kingdom
23	536368	22960	JAM MAKING SET W	6	01-12-2010 08:34	4.25	13047	United Kingdom
24	536368	22913	RED COAT RACK PAR	3	01-12-2010 08:34	4.95	13047	United Kingdom
25	536368	22912	YELLOW COAT RACK	3	01-12-2010 08:34	4.95	13047	United Kingdom
26	536368	22914	BLUE COAT RACK PA	3	01-12-2010 08:34	4.95	13047	United Kingdom
27	536369	21756	BATH BUILDING BLO	3	01-12-2010 08:35	5.95	13047	United Kingdom

File Edit Selection View Go Run Terminal Help ML Project

EXPLORER

ML PROJECT
- Data-Driven Customer Segmen...
- Mall_Customers.csv
- Online Retail.xlsx

OUTLINE
TIMELINE

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])

+ Code + Markdown | Run All | Clear All Outputs | Outline ... Select Kernel

```
#Import important libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.cm as cm
from mpl_toolkits.mplot3d import Axes3D
import seaborn as sns
import datetime as dt

from sklearn.preprocessing import PowerTransformer, StandardScaler

from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score
```

[1] Python

```
/usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use the functions in the public API
  import pandas.util.testing as tm
```

Data Understanding

```
df = pd.read_excel('Online Retail.csv',  parse_dates=['InvoiceDate'])
df.head()
```

Python

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

⊗ 10 ⚠ 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke Search ENG IN 20:20 21-02-2025

File Edit Selection View Go Run Terminal Help

ML Project

EXPLORER

ML PROJECT

Data-Driven Customer Segmen...

Mall_Customers.csv

Online Retail.xlsx

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])

+ Code + Markdown | Run All Clear All Outputs | Outline ... Select Kernel

Data Understanding

```
df = pd.read_excel('Online Retail.csv',  parse_dates=['InvoiceDate'])
df.head()
```

Python

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

```
df.info()
```

Python

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    541909 non-null  object
 1   StockCode    541909 non-null  object
 2   Description  540455 non-null  object
 3   Quantity     541909 non-null  int64
 4   InvoiceDate  541909 non-null  datetime64[ns]
 5   UnitPrice    541909 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      541909 non-null  object
```

OUTLINE

TIMELINE

10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link

Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke

Search

ENG IN

20:20 21-02-2025

File Edit Selection View Go Run Terminal Help ML Project

EXPLORER

ML PROJECT

Data-Driven Customer Segmen...

Mall_Customers.csv

Online Retail.xlsx

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])

+ Code + Markdown | Run All Clear All Outputs | Outline ... Select Kernel

```
df.shape
```

[4] Python

```
(541909, 8)
```

```
df.describe().round(2)
```

[5] Python

	Quantity	UnitPrice	CustomerID
count	541909.00	541909.00	406829.00
mean	9.55	4.61	15287.69
std	218.08	96.76	1713.60
min	-80995.00	-11062.06	12346.00
25%	1.00	1.25	13953.00
50%	3.00	2.08	15152.00
75%	10.00	4.13	16791.00
max	80995.00	38970.00	18287.00

Data Cleansing

```
#Ensure that the data type in the InvoiceNo column is in the form of a string.
df['InvoiceNo'] = df['InvoiceNo'].astype('str')
```

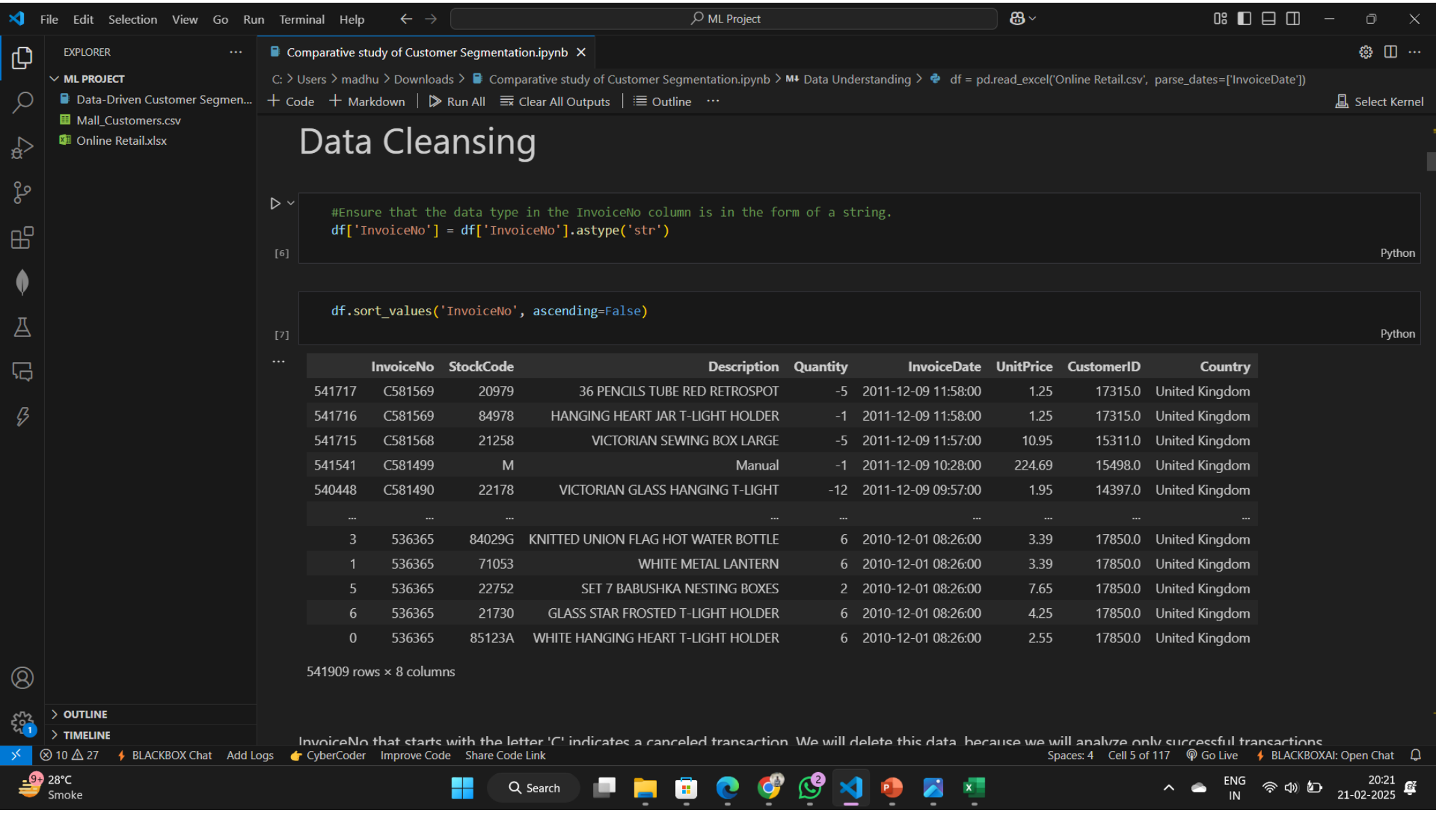
[6] Python

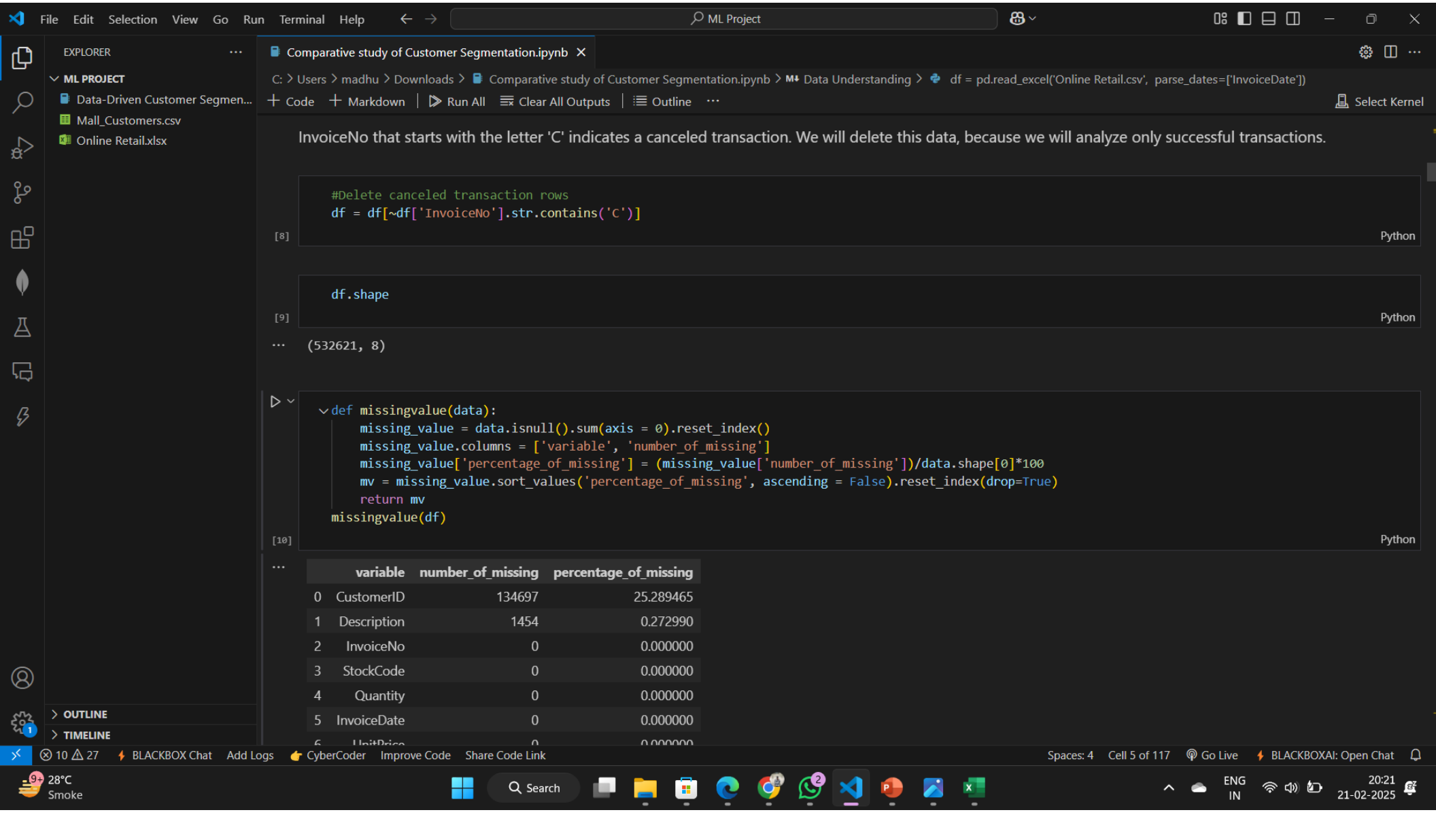
OUTLINE

TIMELINE

10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke Search ENG IN 20:20 21-02-2025


File Edit Selection View Go Run Terminal Help
ML Project
EXPLORER
ML PROJECT
Data-Driven Customer Segmen...
Mall_Customers.csv
Online Retail.xlsx
Comparative study of Customer Segmentation.ipynb
C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])
Code Markdown Run All Clear All Outputs Outline
Select Kernel
Data Cleansing
#Ensure that the data type in the InvoiceNo column is in the form of a string.
df['InvoiceNo'] = df['InvoiceNo'].astype('str')
[6]
Python
df.sort_values('InvoiceNo', ascending=False)
[7]
Python
InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
541717 C581569 20979 36 PENCILS TUBE RED RETROSPOT -5 2011-12-09 11:58:00 1.25 17315.0 United Kingdom
541716 C581569 84978 HANGING HEART JAR T-LIGHT HOLDER -1 2011-12-09 11:58:00 1.25 17315.0 United Kingdom
541715 C581568 21258 VICTORIAN SEWING BOX LARGE -5 2011-12-09 11:57:00 10.95 15311.0 United Kingdom
541541 C581499 M Manual -1 2011-12-09 10:28:00 224.69 15498.0 United Kingdom
540448 C581490 22178 VICTORIAN GLASS HANGING T-LIGHT -12 2011-12-09 09:57:00 1.95 14397.0 United Kingdom
... ...
3 536365 84029G KNITTED UNION FLAG HOT WATER BOTTLE 6 2010-12-01 08:26:00 3.39 17850.0 United Kingdom
1 536365 71053 WHITE METAL LANTERN 6 2010-12-01 08:26:00 3.39 17850.0 United Kingdom
5 536365 22752 SET 7 BABUSHKA NESTING BOXES 2 2010-12-01 08:26:00 7.65 17850.0 United Kingdom
6 536365 21730 GLASS STAR FROSTED T-LIGHT HOLDER 6 2010-12-01 08:26:00 4.25 17850.0 United Kingdom
0 536365 85123A WHITE HANGING HEART T-LIGHT HOLDER 6 2010-12-01 08:26:00 2.55 17850.0 United Kingdom
541909 rows × 8 columns
InvoiceNo that starts with the letter 'C' indicates a canceled transaction. We will delete this data, because we will analyze only successful transactions
OUTLINE
TIMELINE
10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link
Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat
28°C Smoke
Search
ENG IN
20:21 21-02-2025


File Edit Selection View Go Run Terminal Help
ML Project
EXPLORER
ML PROJECT
Data-Driven Customer Segmen...
Mall_Customers.csv
Online Retail.xlsx
Comparative study of Customer Segmentation.ipynb
C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])
Code Markdown Run All Clear All Outputs Outline
Select Kernel
InvoiceNo that starts with the letter 'C' indicates a canceled transaction. We will delete this data, because we will analyze only successful transactions.
#Delete canceled transaction rows
df = df[~df['InvoiceNo'].str.contains('C')]
[8]
Python
df.shape
[9]
Python
(532621, 8)
def missingvalue(data):
missing_value = data.isnull().sum(axis = 0).reset_index()
missing_value.columns = ['variable', 'number_of_missing']
missing_value['percentage_of_missing'] = (missing_value['number_of_missing'])/data.shape[0]*100
mv = missing_value.sort_values('percentage_of_missing', ascending = False).reset_index(drop=True)
return mv
missingvalue(df)
[10]
Python
variable number_of_missing percentage_of_missing
0 CustomerID 134697 25.289465
1 Description 1454 0.272990
2 InvoiceNo 0 0.000000
3 StockCode 0 0.000000
4 Quantity 0 0.000000
5 InvoiceDate 0 0.000000
OUTLINE
TIMELINE
10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link
Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat
28°C Smoke
Search
ENG IN
20:21 21-02-2025

File Edit Selection View Go Run Terminal Help ML Project

EXPLORER

ML PROJECT
- Data-Driven Customer Segmen...
- Mall_Customers.csv
- Online Retail.xlsx

OUTLINE
TIMELINE

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])

+ Code + Markdown | Run All | Clear All Outputs | Outline ... Select Kernel

Additionally, CustomerID and Description can not be filled because it will affect the analysis later.

[11]
```
df = df.dropna(subset = ['CustomerID', 'Description'])
```
Python

[12]
```
missingvalue(df)
```
Python

	variable	number_of_missing	percentage_of_missing
0	InvoiceNo	0	0.0
1	StockCode	0	0.0
2	Description	0	0.0
3	Quantity	0	0.0
4	InvoiceDate	0	0.0
5	UnitPrice	0	0.0
6	CustomerID	0	0.0
7	Country	0	0.0

[13]
```
#Remove spaces at the beginning or end of sentences
df['Description'] = df['Description'].str.strip()
```
Python

[14]
```
df.shape
```
Python

(397924, 8)

10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke Search ENG IN 20:21 21-02-2025

File Edit Selection View Go Run Terminal Help ML Project

EXPLORER

ML PROJECT
- Data-Driven Customer Segmen...
- Mall_Customers.csv
- Online Retail.xlsx

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])

+ Code + Markdown | Run All | Clear All Outputs | Outline ... | Select Kernel

[14]
```
df.shape
```
Python

```
(397924, 8)
```

[15]
```
df.info()
df['CustomerID'] = df['CustomerID'].astype('int')
```
Python

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 397924 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    397924 non-null  object
 1   StockCode    397924 non-null  object
 2   Description  397924 non-null  object
 3   Quantity     397924 non-null  int64
 4   InvoiceDate  397924 non-null  datetime64[ns]
 5   UnitPrice    397924 non-null  float64
 6   CustomerID   397924 non-null  float64
 7   Country      397924 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 27.3+ MB
```

[16]
```
df.describe()
```
Python

	Quantity	UnitPrice	CustomerID
count	397924.000000	397924.000000	397924.000000

OUTLINE
TIMELINE

10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link

Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke

Search

ENG IN 20:21 21-02-2025

File Edit Selection View Go Run Terminal Help — ML Project

EXPLORER

ML PROJECT
- Data-Driven Customer Segmen...
- Mall_Customers.csv
- Online Retail.xlsx

OUTLINE
TIMELINE

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])

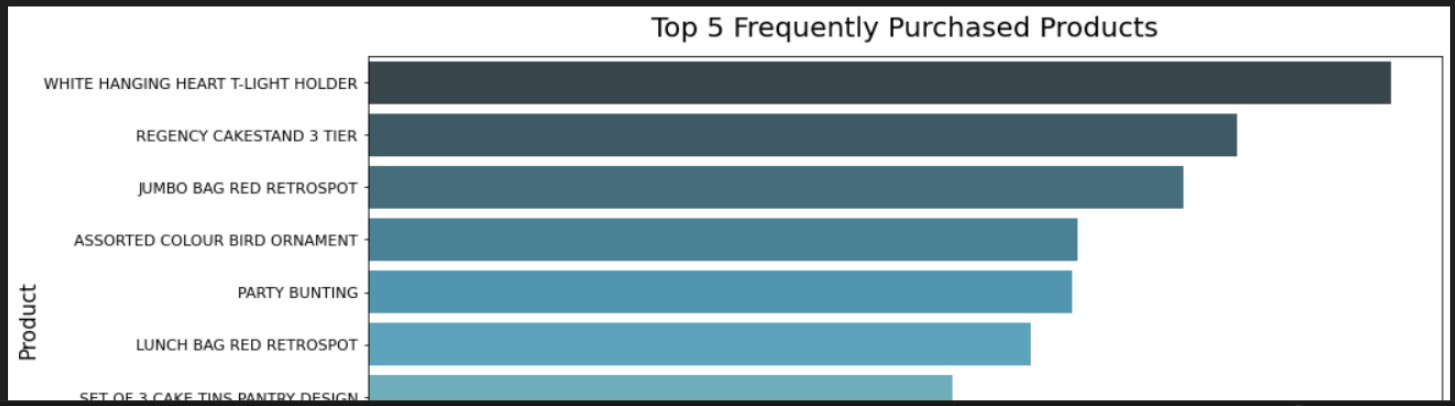
+ Code + Markdown | Run All | Clear All Outputs | Outline ... | Select Kernel

Exploratory Data Analysis

What products do customers buy the most?

```
plt.figure(figsize = (14,7))
ax = sns.countplot(y = df['Description'],
                   order = df["Description"].value_counts().iloc[:10].index,
                   palette = "GnBu_d")
ax.set_title("Top 5 Frequently Purchased Products", size = 20, pad=15)
ax.set_xlabel("Count", size = 15)
ax.set_ylabel("Product", size = 15)
ax.xaxis.set_tick_params(labelsize=11)
ax.yaxis.set_tick_params(labelsize=11)
plt.show()
```

[17] Python



10 ⚠ 27 | BLACKBOX Chat | Add Logs | CyberCoder | Improve Code | Share Code Link | Spaces: 4 | Cell 5 of 117 | Go Live | BLACKBOXAI: Open Chat

28°C Smoke | Search | ENG IN | 20:21 21-02-2025

File Edit Selection View Go Run Terminal Help ML Project

EXPLORER

ML PROJECT

Data-Driven Customer Segmen...

Mall_Customers.csv

Online Retail.xlsx

Comparative study of Customer Segmentation.ipynb

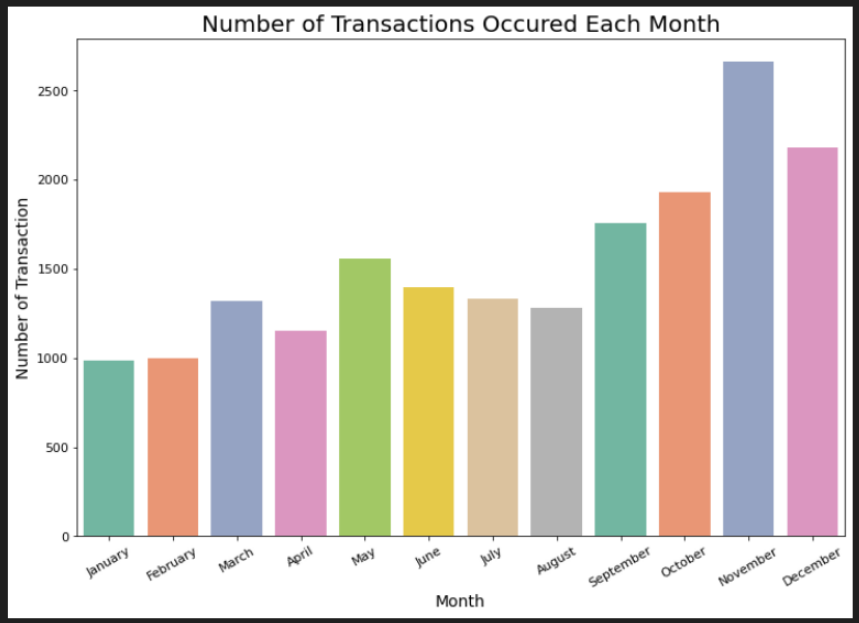
C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])

+ Code + Markdown | Run All | Clear All Outputs | Outline ... Select Kernel

```
order_month = df.groupby(df['InvoiceDate'].dt.month)['InvoiceNo'].nunique()

plt.figure(figsize = (12,8))
ax = sns.barplot(x = 'InvoiceDate', y = 'InvoiceNo', data = order_month.reset_index(), palette = "Set2")
ax.set_title('Number of Transactions Occured Each Month', size = 20)
ax.set_xlabel('Month', size = 14)
ax.set_ylabel('Number of Transaction', size = 14)
ax.xaxis.set_tick_params(labelsize=11)
ax.yaxis.set_tick_params(labelsize=11)
ax.set_xticklabels(['January','February','March','April','May','June','July','August','September','October','November','December'], rotation = 30)
plt.show()
```

[18] Python



OUTLINE

TIMELINE

10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke Search ENG IN 20:21 21-02-2025

File Edit Selection View Go Run Terminal Help ML Project

EXPLORER

ML PROJECT

Data-Driven Customer Segmen...

Mall_Customers.csv

Online Retail.xlsx

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Data Understanding > df = pd.read_excel('Online Retail.csv', parse_dates=['InvoiceDate'])

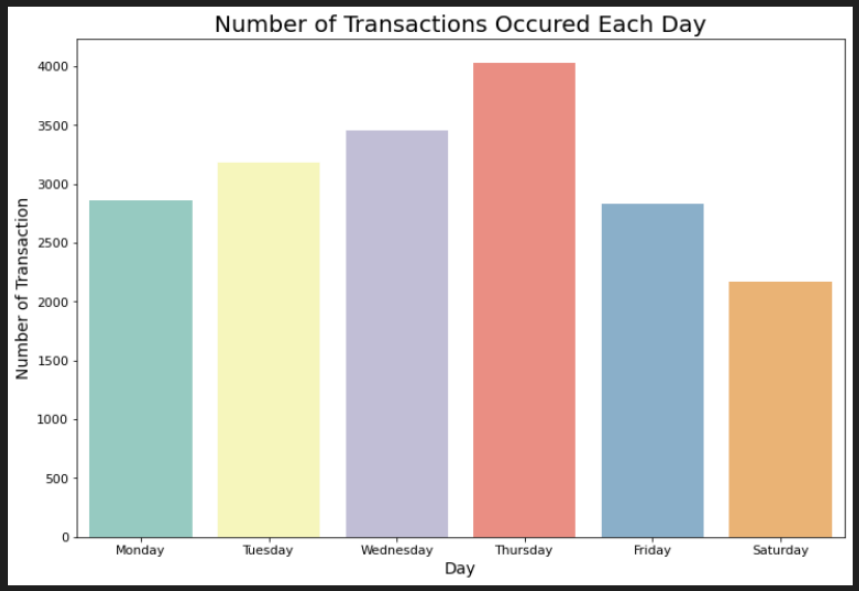
+ Code + Markdown | Run All | Clear All Outputs | Outline ... Select Kernel

How many orders (per day)?

```
order_day = df.groupby(df['InvoiceDate'].dt.dayofweek)['InvoiceNo'].nunique()

plt.figure(figsize = (12,8))
ax = sns.barplot(x = 'InvoiceDate', y = 'InvoiceNo', data = order_day.reset_index(), palette = "Set3")
ax.set_title('Number of Transactions Occured Each Day', size = 20)
ax.set_xlabel('Day', size = 14)
ax.set_ylabel('Number of Transaction', size = 14)
ax.xaxis.set_tick_params(labelsize=11)
ax.yaxis.set_tick_params(labelsize=11)
ax.set_xticklabels(['Monday','Tuesday','Wednesday','Thursday','Friday','Saturday'])
plt.show()
```

[19] Python



OUTLINE

TIMELINE

10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link Spaces: 4 Cell 5 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke Search ENG IN 20:21 21-02-2025

File Edit Selection View Go Run Terminal Help — ML Project

EXPLORER

ML PROJECT
- Data-Driven Customer Segmen...
- Mall_Customers.csv
- Online Retail.xlsx

Comparative study of Customer Segmentation.ipynb

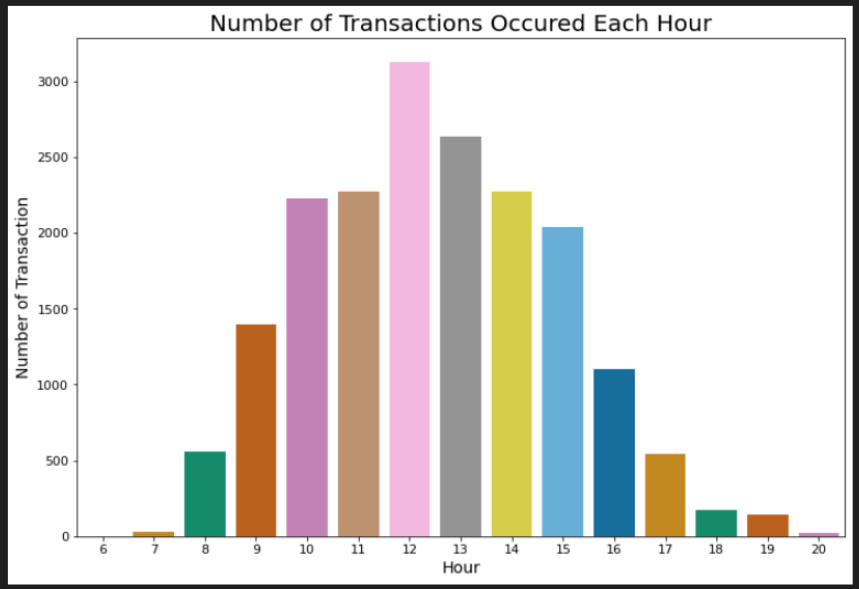
C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Exploratory Data Analysis > How many orders (per hour)? > order_hour = df.groupby(df['InvoiceDate'].dt.hour)['InvoiceNo'].nunique()

+ Code + Markdown | Run All | Clear All Outputs | Outline ... Select Kernel

```
order_hour = df.groupby(df['InvoiceDate'].dt.hour)['InvoiceNo'].nunique()

plt.figure(figsize = (12,8))
ax = sns.barplot(x = 'InvoiceDate', y = 'InvoiceNo', data = order_hour.reset_index(), palette = "colorblind")
ax.set_title('Number of Transactions Occured Each Hour', size = 20)
ax.set_xlabel('Hour', size = 14)
ax.set_ylabel('Number of Transaction', size = 14)
ax.xaxis.set_tick_params(labelsize=11)
ax.yaxis.set_tick_params(labelsize=11)
plt.show()
```

[20] Python



OUTLINE
TIMELINE

+ Code + Markdown

10 ⚠ 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link — Spaces: 4 Cell 34 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke — Search — ENG IN — 20:32 21-02-2025

File Edit Selection View Go Run Terminal Help ML Project

EXPLORER
ML PROJECT
Data-Driven Customer Segmen...
Mall_Customers.csv
Online Retail.xlsx

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Exploratory Data Analysis > How many orders (per hour)? > order_hour = df.groupby(df['InvoiceDate'].dt.hour)['InvoiceNo'].nunique()

+ Code + Markdown | Run All Clear All Outputs | Outline ... Select Kernel

Who are the 5 customers who spend the most money on Online Retail?

```
df['TotalPrice'] = df['Quantity']*df['UnitPrice']
df.head()
```

[21] Python

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	TotalPrice
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850	United Kingdom	15.30
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850	United Kingdom	22.00
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34

```
price_cust = pd.pivot_table(df, index='CustomerID', values='TotalPrice', aggfunc=np.sum)
print('The following are 5 customers who spend the most money on Online Retail:')
price_cust.sort_values('TotalPrice', ascending=False)[:5]
```

[22] Python

```
The following are 5 customers who spend the most money on Online Retail:
```

CustomerID	TotalPrice
14646	280206.02
18102	259657.30
17450	194550.79
16446	168472.50
14911	143825.06

Customer with ID 14646 is the customer who spent the most money in online retail, which is $280,206.02

Who are the five most frequent customer shopping in Online Retail?

OUTLINE
TIMELINE

10 ⚠ 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link Spaces: 4 Cell 34 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke Search ENG IN 20:32 21-02-2025

File Edit Selection View Go Run Terminal Help ML Project

EXPLORER

ML PROJECT

Data-Driven Customer Segmen...

Mall_Customers.csv

Online Retail.xlsx

Comparative study of Customer Segmentation.ipynb

C: > Users > madhu > Downloads > Comparative study of Customer Segmentation.ipynb > Exploratory Data Analysis > How many orders (per hour)? > order_hour = df.groupby(df['InvoiceDate'].dt.hour)['InvoiceNo'].nunique()

+ Code + Markdown | Run All Clear All Outputs | Outline ... Select Kernel

Who are the five most frequent customer shopping in Online Retail?

```
order_cust = pd.pivot_table(df, index='CustomerID', values='InvoiceNo', aggfunc=pd.Series.nunique)
print('The following are the 5 customers who most frequently shop at Online Retail:')
order_cust.sort_values('InvoiceNo', ascending=False)[:5]
```

[23] Python

```
The following are the 5 customers who most frequently shop at Online Retail:
```

CustomerID	InvoiceNo
12748	210
14911	201
17841	124
13089	97
14606	93

Customer with ID 12748 is the most frequent customer shopping in online retail, with 210 transactions.
Customers with ID 14911 are customers who frequently shop and spend a lot of money at online retail.

Time Cohorts

First of all, we will retrieve the InvoiceMonth data.

```
def get_month(x):
    return dt.datetime(x.year, x.month, 1)
df['InvoiceMonth'] = df['InvoiceDate'].apply(get_month)
```

[24] Python

```
df.head()
```

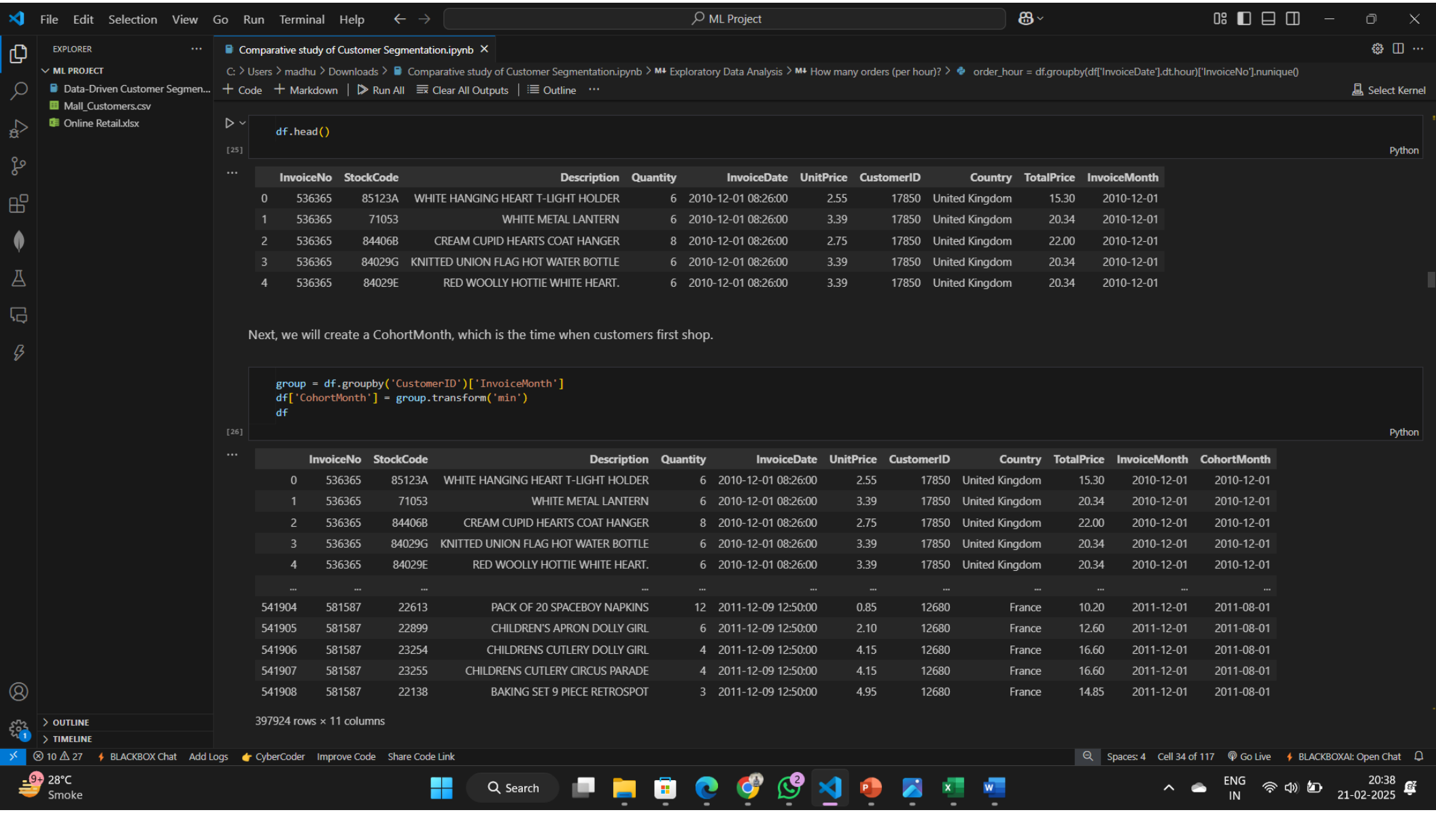
[25] Python

OUTLINE

TIMELINE

10 27 BLACKBOX Chat Add Logs CyberCoder Improve Code Share Code Link Spaces: 4 Cell 34 of 117 Go Live BLACKBOXAI: Open Chat

28°C Smoke Search ENG IN 20:38 21-02-2025



	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	TotalPrice	InvoiceMonth
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850	United Kingdom	15.30	2010-12-01
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34	2010-12-01
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850	United Kingdom	22.00	2010-12-01
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34	2010-12-01
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34	2010-12-01

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	TotalPrice	InvoiceMonth	CohortMonth
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850	United Kingdom	15.30	2010-12-01	2010-12-01
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34	2010-12-01	2010-12-01
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850	United Kingdom	22.00	2010-12-01	2010-12-01
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34	2010-12-01	2010-12-01
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850	United Kingdom	20.34	2010-12-01	2010-12-01
...	...	...	...	...	...	...	...	...	...	...	...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	2011-12-09 12:50:00	0.85	12680	France	10.20	2011-12-01	2011-08-01
541905	581587	22899	CHILDREN'S APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680	France	12.60	2011-12-01	2011-08-01
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680	France	16.60	2011-12-01	2011-08-01
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680	France	16.60	2011-12-01	2011-08-01
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680	France	14.85	2011-12-01	2011-08-01

Article No.	Title	Reference	Introduction	Conclusion
1	Data-Driven Customer Segmentation: Advancing Precision Marketing through Analytics and Machine Learning Techniques	Ur Rahaman S, Kumar S, Puchakayala PRA. J Artif Intell Mach Learn Data Sci. 2021;1(1):1356-1362.	Critiques traditional segmentation techniques, emphasizing data analytics and machine learning for real-time behavioral, transactional, and interaction data.	Machine learning techniques outperform traditional segmentation, uncover hidden behavior patterns, and improve personalized marketing strategies.
2	UNICON: A Unified Framework for Behavior-Based Consumer Segmentation in E-Commerce	Dibak M, Vlasov V, Karessli N, et al. arXiv preprint arXiv:2309.13068. 2023.	Introduces UNICON, a deep learning framework for consumer behavior-based segmentation in fashion e-commerce.	UNICON effectively bridges traditional segmentation with deep learning, improving hybrid recommender systems and customer experience.
3	Predictive Analytics for Customer Behavior Modeling in E-Commerce Using Classification Algorithms	Tan L, Liu X, Zhang H. IEEE Trans Ind Inform. 2018;14(6):2442-2451.	Examines clickstream data to predict customer conversion rates using classification models.	Predictive analytics significantly enhances conversion predictions, providing actionable insights for targeted marketing.
4	Customer Segmentation and Predictive Analytics: A Case Study of E-Commerce	Li H, Zhang L. J Bus Res. 2017;64(5):789-803.	Explores integrating transactional and behavioral data for precise customer segmentation in e-commerce.	Combining historical and real-time data enhances targeted marketing, improving conversion rates and retention.
5	Improving Customer Retention through Predictive Modeling: The Role of Machine Learning	Kumar A, Shankar R. E-Commer Bus Intell. 2018;6(4):98-112.	Investigates machine learning models for predicting customer churn and designing retention strategies.	Predictive models help forecast churn, enabling personalized retention strategies and improved customer lifetime value.
6	RFM and CLV: Using Iso-Value Curves for Customer Base Analysis	Fader PS, Hardie BG, Lee KL. J Mark Res. 2005;42(4):415-430.	Integrates Recency, Frequency, and Monetary (RFM) analysis with Customer Lifetime Value (CLV) estimation.	Combining RFM and CLV enhances segmentation accuracy and long-term profitability predictions.
7	Clustering Algorithms in Customer Segmentation: A Review	Yang J, Liu Z. IEEE Trans Knowl Data Eng. 2018;30(5):964-977.	Reviews clustering techniques for customer segmentation, including K-means, hierarchical, and density-based methods.	No single best clustering method; selection should be data-specific for better segmentation accuracy.

8	A Predictive Model for Customer Segmentation Using Machine Learning Techniques	Bhat SSS, Kulkarni PD, Murthy SVS. Proc IEEE Int Conf Comput Intell Comput Res; 2014. p. 420-424.	Proposes a hybrid predictive model using supervised and unsupervised machine learning techniques.	Hybrid models improve segmentation accuracy, allowing businesses to create better customer marketing strategies.
9	Customer Segmentation Using Real Transactional Data in E-Commerce Platform: A Case of Online Fashion Bags Shop	Yan ZZ, Zhao Y. Proc Int Conf Electron Bus; 2021 Dec 3-7; Nanjing, China. p. 90-99.	Applies Fuzzy C-Means clustering and probabilistic neural networks for customer segmentation.	Using real transactional data improves segmentation accuracy and predictive marketing strategies.
10	Optimizing Predictive Models for Customer Segmentation in E-Commerce: A Data Science Approach	Ur Rahaman S. Int J Innov Res Creat Technol. 2018;4(4).	Benchmarks machine learning algorithms for optimizing customer segmentation models.	Ensemble and hybrid machine learning models achieve superior segmentation accuracy.
11	An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market	John JM, Shobayo O, Ogunleye B. arXiv preprint arXiv:2402.04103. 2024.	Examines clustering algorithms for customer segmentation in UK retail, using RFM model.	Gaussian Mixture Model achieved highest segmentation accuracy with a Silhouette Score of 0.80.
12	A Hybrid Statistical-Machine Learning Approach for Analyzing Online Customer Behavior	Alizamir S, Bandara K, Eshragh A, Iravani F. arXiv preprint arXiv:2212.02255. 2022.	Combines statistical methods and machine learning to analyze online customer behavior in JD.com.	Finds delivery speed significantly influences order quantity, while price is the main decision factor.
13	Customer Segmentation Using AI-Based Behavioral Analytics in E-Commerce	Sharma P, Gupta R, Nair B. J Retail Market Analytics. 2023;5(3):205-217.	Explores AI-driven behavioral analytics for identifying e-commerce customer segments.	AI models outperform traditional segmentation, revealing more dynamic consumer patterns.
14	E-Commerce Customer Segmentation: A Data Mining Approach	Chowdhury K, Hassan M. Data Sci J. 2022;19(2):98-112.	Applies data mining techniques like Apriori algorithm and decision trees for segmentation.	Finds hybrid models of clustering and classification provide most actionable customer insights.
15	Advanced Personalization Strategies Using Real-Time Customer Segmentation	Zhang X, Lee C, Park H. J Big Data Anal. 2021;8(1):77-91.	Develops real-time segmentation methods using neural networks and deep learning.	Real-time segmentation enables superior personalization, boosting sales and retention.

16	Using Big Data for E-Commerce Customer Segmentation: A Review	Wei Y, Chen B. J Big Data. 2023;10(5):32-49.	Reviews applications of big data analytics in customer segmentation strategies.	Finds clustering combined with machine learning improves targeting and marketing ROI.
17	Impact of Customer Segmentation on Predictive Marketing in Online Retail	Nielsen M, Hoffmann P. Int J Mark Res. 2022;14(4):312-326.	Studies predictive analytics in marketing strategies based on customer segmentation.	Data-driven segmentation improves ROI and customer satisfaction significantly.
18	Machine Learning for Customer Segmentation in Fashion Retail	Kim J, Lee H. Comput Ind Eng. 2023;12(1):54-66.	Explores machine learning applications in segmenting fashion retail customers.	Finds deep learning-based segmentation methods improve accuracy and sales conversions.
19	Clustering Consumer Behaviors for E-Commerce Optimization	Anderson P, Brown T. IEEE Trans Cybern. 2023;50(7):1123-1135.	Uses clustering algorithms to analyze and categorize consumer behaviors in e-commerce.	Hybrid clustering approaches are most effective in detecting emerging market trends.
20	AI-Powered Customer Segmentation for Hyper-Personalized Marketing	Gomez R, Patel S. J Artif Intell Appl. 2024;9(2):100-115.	Develops AI-based segmentation framework integrating sentiment analysis and purchase behaviors.	AI-powered segmentation significantly enhances personalization, engagement, and conversion rates.

Conclusion: Embracing the Future of Marketing

1. Data-driven customer segmentation is a powerful tool for enhancing marketing strategies, driving engagement, and boosting revenue.

2. This project has highlighted the immense potential of AI and ML in transforming the marketing landscape.

3. By integrating AI-driven segmentation into their strategies, businesses can unlock new opportunities for growth and success in the modern marketplace.