Exploring E-commerce Data:
A Practical Analysis

A Micro Project Report

Submitted by

RAJESH KANNA R.
Reg.no: 99220041074

B.Tech – Computer Science Engineering,
Data Science

Kalasalingam Academy of Research and Education

(Deemed to be University)

Anand Nagar, Krishnankoil - 626 126

March-2024

KALASALINGAM
ACADEMY OF RESEARCH AND EDUCATION
(DEEMED TO BE UNIVERSITY)
Under sec. 3 of UGC Act 1956. Accredited by NAAC with "A" Grade

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Bonafide record of the work done by RAJESH KANNA R. - 99220041074 in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Specialization of the Computer Science and Engineering, during the Academic Year [Even] Semester (2023-24).

Dr.G.Nagarajan
Project Guide
Associate Professor
Dept. CSE
Kalasalingam Academy of
Research and Education
Krishnankoil – 626126

Mr.Gnana Kumar
Faculty incharge
Professor
Dept. CSE
Kalasalingam Academy of
Research and Education
Krishnankoil - 626126

Mr.M.Jafer Sathick Ali
Evaluator
Assistant Professor
Dept. CSE
Kalasalingam Academy of
Research and Education
Krishnankoil - 626126

Abstract

This project delves into the analysis of an e-commerce dataset using the Pandas library in Python, conducted within a Jupyter notebook environment. The dataset includes information such as purchase prices, credit card providers, job titles, and languages of customers. The analysis begins with basic exploratory tasks, such as displaying the top and last rows, checking for null values, and summarizing the dataset's structure.

Further analysis includes identifying trends such as the highest and lowest purchase prices, the average purchase price, and the distribution of purchases by language and job title. Intermediate-level analysis involves filtering data to find specific patterns, such as customers with Mastercard as their credit card provider who made purchases above a certain threshold or those with credit cards expiring in a particular year.

Data visualization plays a crucial role in this analysis, with various plots illustrating purchase counts, purchase prices, and company purchases. Scatter plots highlight purchase prices based on specific job titles, while bar charts showcase average purchase prices by language and browser usage. These visualizations offer insights into customer behavior and preferences within the e-commerce platform.

This project demonstrates the versatility of Pandas in analyzing and visualizing complex datasets, providing valuable insights for e-commerce businesses to understand their customer base better and make informed decisions.

Contents

Introduction

The proliferation of e-commerce platforms has led to an abundance of data that provides valuable insights into consumer behavior and market trends. In this project, we analyze the 'E-Commerce Purchase' dataset, downloaded from Kaggle, using the powerful data manipulation and analysis library, Pandas, in Python. This dataset contains information on e-commerce purchases, including details such as purchase prices, credit card providers, job titles, and languages of customers etc.

Our objective is to conduct a comprehensive analysis of this dataset, exploring various aspects such as purchase trends, customer demographics, and popular products. By leveraging the capabilities of Pandas, we aim to extract meaningful insights that can help e-commerce businesses make informed decisions and enhance their understanding of customer preferences.

Through this analysis, we aim to demonstrate the effectiveness of Pandas in handling and analyzing real-world datasets, highlighting its utility in extracting valuable insights from complex data. The findings of this analysis have the potential to inform marketing strategies, product offerings, and customer engagement initiatives, ultimately contributing to the growth and success of e-commerce businesses.

What is Pandas?

Pandas is an open-source data manipulation and analysis library for Python. It provides easy-to-use data structures and functions designed to make working with structured data fast, easy, and expressive. Pandas is built on top of NumPy, another Python library that provides support for large, multi-dimensional arrays and matrices, making it a powerful tool for data analysis and manipulation.

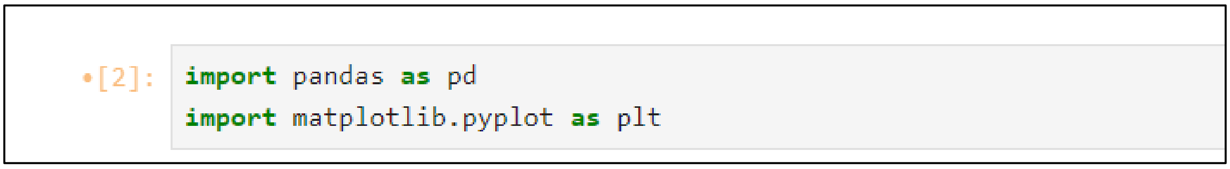
The primary data structures in Pandas are Series and DataFrame. A Series is a one-dimensional array-like object that can hold various data types, such as integers, strings, or floating-point numbers. A DataFrame is a two-dimensional, size-mutable, and heterogeneous tabular data structure with labeled axes (rows and columns).

Pandas provides a wide range of functions for reading and writing data, data cleaning, reshaping, merging, slicing, indexing, and aggregating data. It also integrates seamlessly with other libraries in the Python ecosystem, such as Matplotlib for data visualization and scikit-learn for machine learning tasks.

In this project, we leverage the power of Pandas to explore and analyze the 'E-Commerce Purchase' dataset, demonstrating its capabilities in handling and analyzing real-world data efficiently.

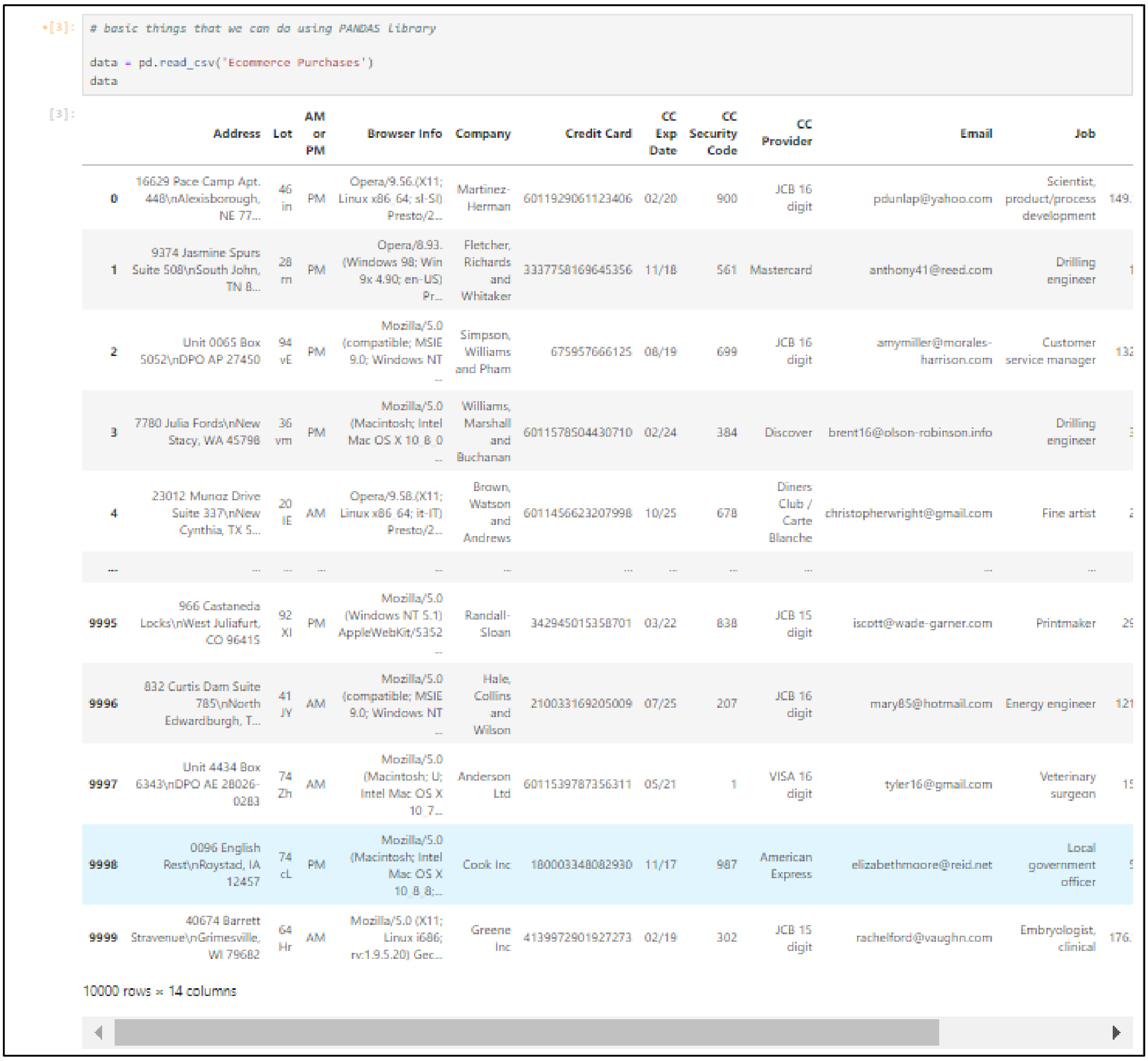
Chapter 1

• Importing Libraries



```
•[2]: import pandas as pd
      import matplotlib.pyplot as plt
```

• Reading the whole Dataset



```
•[3]: # basic things that we can do using PANDAS Library

      data = pd.read_csv('Ecommerce Purchases')
      data
```

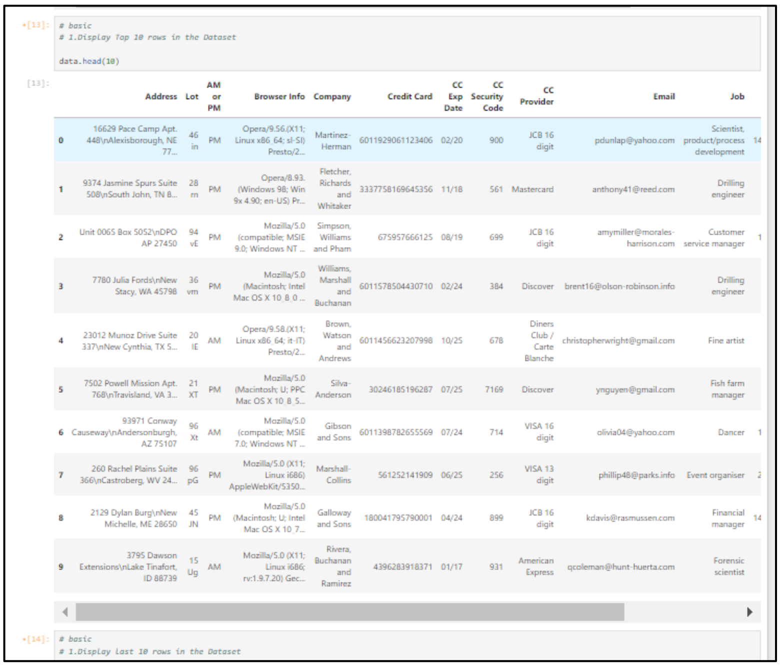
[3]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email	Job	
0	16629 Pace Camp Apt. 448\nAlexisborough, NE 77...	46 in	PM	Opera/9.56.(X11; Linux x86_64; sl-SI) Presto/2...	Martinez-Herman	6011929061123406	02/20	900	JCB 16 digit	pdunlap@yahoo.com	Scientist, product/process development	149.
1	9374 Jasmine Spurs Suite 508\nSouth John, TN 8...	28 rn	PM	Opera/8.93.(Windows 98; Win 9x 4.90; en-US) Pr...	Fletcher, Richards and Whitaker	3337758169645356	11/18	561	Mastercard	anthony41@reed.com	Drilling engineer	1
2	Unit 0065 Box 5052\nDPO AP 27450	94 vE	PM	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Simpson, Williams and Pham	675957666125	08/19	699	JCB 16 digit	amymiller@morales-harrison.com	Customer service manager	132
3	7780 Julia Fords\nNew Stacy, WA 45798	36 vm	PM	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8_0 ...	Williams, Marshall and Buchanan	6011578504430710	02/24	384	Discover	brent16@olson-robinson.info	Drilling engineer	3
4	23012 Munoz Drive Suite 337\nNew Cynthia, TX 5...	20 IE	AM	Opera/9.58.(X11; Linux x86_64; it-IT) Presto/2...	Brown, Watson and Andrews	6011456623207998	10/25	678	Diners Club / Carte Blanche	christopherwright@gmail.com	Fine artist	2
...	...	...	...	...	...	...	...	...	...	...	...	
9995	966 Castaneda Locks\nWest Juliafurt, CO 96415	92 XI	PM	Mozilla/5.0 (Windows NT 5.1) AppleWebKit/5352 ...	Randall-Sloan	342945015358701	03/22	838	JCB 15 digit	iscott@wade-garner.com	Printmaker	29
9996	832 Curtis Dam Suite 785\nNorth Edwardburgh, T...	41 JY	AM	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Hale, Collins and Wilson	210033169205009	07/25	207	JCB 16 digit	mary85@hotmail.com	Energy engineer	121
9997	Unit 4434 Box 6343\nDPO AE 28026-0283	74 Zh	AM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_7...	Anderson Ltd	6011539787356311	05/21	1	VISA 16 digit	tyler16@gmail.com	Veterinary surgeon	15
9998	0096 English Rest\nRoystad, IA 12457	74 cL	PM	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8_8;...	Cook Inc	180003348082930	11/17	987	American Express	elizabethmoore@reid.net	Local government officer	5
9999	40674 Barrett Stravenue\nGrimesville, WI 79682	64 Hr	AM	Mozilla/5.0 (X11; Linux i686; rv:1.9.5.20) Gec...	Greene Inc	4139972901927273	02/19	302	JCB 15 digit	rachelford@vaughn.com	Embryologist, clinical	176.

10000 rows × 14 columns

Chapter 2 - Basic things

- Reading first 10 column in the Dataset



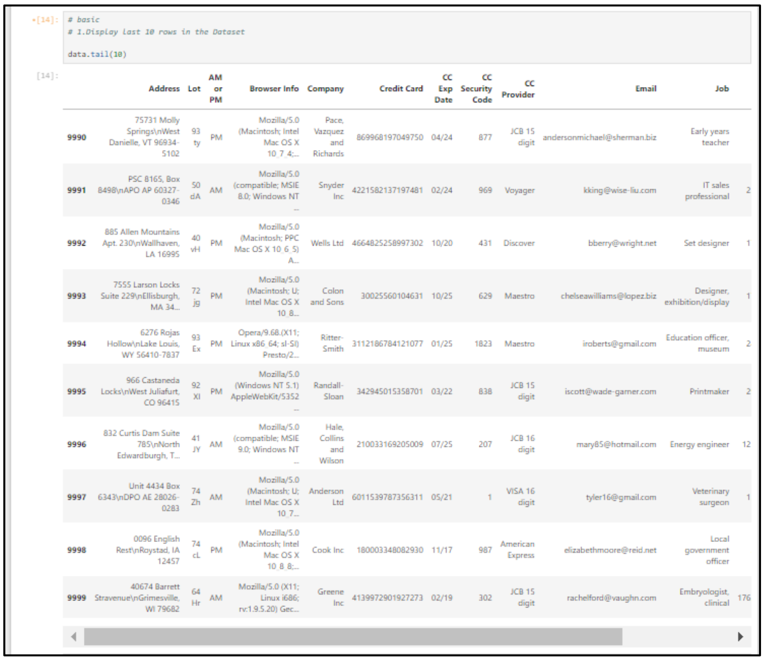
[13]:
```
# basic
# 1.Display Top 10 rows in the Dataset

data.head(10)
```

[13]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email	Job	
0	16629 Pace Camp Apt. 448\nAlexisborough, NE 77...	46 in	PM	Opera/9.56.(X11; Linux x86_64; sl-SI) Presto/2...	Martinez-Herman	6011929061123406	02/20	900	JCB 16 digit	pdunlap@yahoo.com	Scientist, product/process development	14
1	9374 Jasmine Spurs Suite 508\nSouth John, TN 8...	28 rn	PM	Opera/8.93.(Windows 98; Win 9x 4.90; en-US) Pr...	Fletcher, Richards and Whitaker	3337758169645356	11/18	561	Mastercard	anthony41@reed.com	Drilling engineer	
2	Unit 0065 Box 5052\nDPO AP 27450	94 vE	PM	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Simpson, Williams and Pham	675957666125	08/19	699	JCB 16 digit	amymiller@morales-harrison.com	Customer service manager	1
3	7780 Julia Fords\nNew Stacy, WA 45798	36 vm	PM	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8_0 ...	Williams, Marshall and Buchanan	6011578504430710	02/24	384	Discover	brent16@olson-robinson.info	Drilling engineer	
4	23012 Munoz Drive Suite 337\nNew Cynthia, TX 5...	20 IE	AM	Opera/9.58.(X11; Linux x86_64; it-IT) Presto/2...	Brown, Watson and Andrews	6011456623207998	10/25	678	Diners Club / Carte Blanche	christopherwright@gmail.com	Fine artist	
5	7502 Powell Mission Apt. 768\nTravisland, VA 3...	21 XT	PM	Mozilla/5.0 (Macintosh; U; PPC Mac OS X 10_8_5...	Silva-Anderson	30246185196287	07/25	7169	Discover	ynguyen@gmail.com	Fish farm manager	
6	93971 Conway Causeway\nAndersonburgh, AZ 75107	96 Xt	AM	Mozilla/5.0 (compatible; MSIE 7.0; Windows NT ...	Gibson and Sons	6011398782655569	07/24	714	VISA 16 digit	olivia04@yahoo.com	Dancer	1
7	260 Rachel Plains Suite 366\nCastroberg, WV 24...	96 pG	PM	Mozilla/5.0 (X11; Linux i686) AppleWebKit/5350...	Marshall-Collins	561252141909	06/25	256	VISA 13 digit	phillip48@parks.info	Event organiser	2
8	2129 Dylan Burg\nNew Michelle, ME 28650	45 JN	PM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_7...	Galloway and Sons	180041795790001	04/24	899	JCB 16 digit	kdavis@rasmussen.com	Financial manager	14
9	3795 Dawson Extensions\nLake Tinafort, ID 88739	15 Ug	AM	Mozilla/5.0 (X11; Linux i686; rv:1.9.7.20) Gec...	Rivera, Buchanan and Ramirez	4396283918371	01/17	931	American Express	qcoleman@hunt-huerta.com	Forensic scientist	

[14]:
```
# basic
# 1.Display last 10 rows in the Dataset
```

- Reading the last 10 columns in the Dataset



[14]:
```
# basic
# 1.Display last 10 rows in the Dataset

data.tail(10)
```

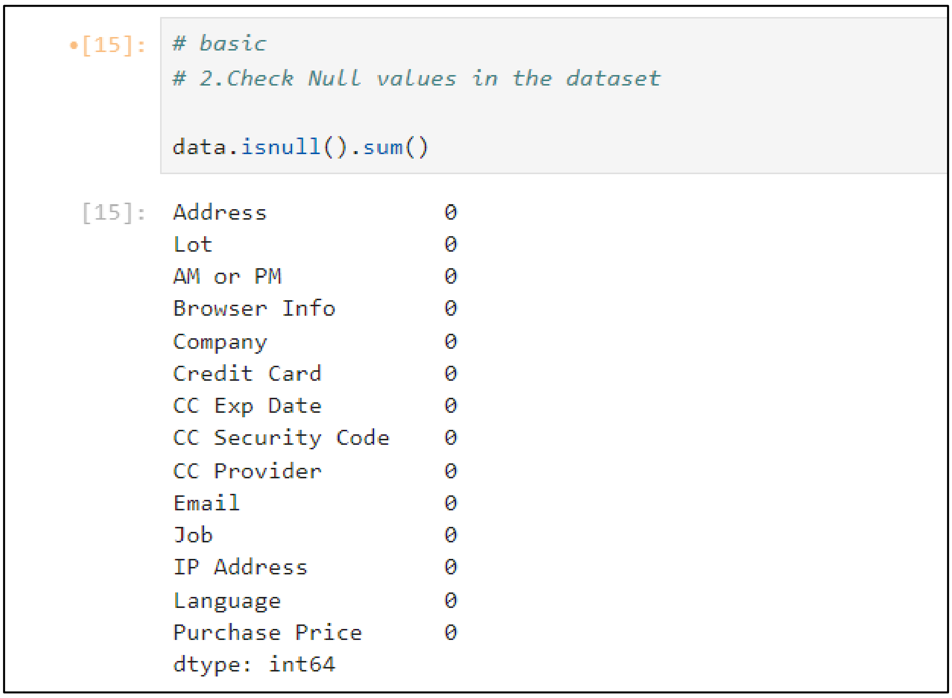
[14]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email	Job	
9990	75731 Molly Springs\nWest Danielle, VT 96934-5102	93 ty	PM	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4;...	Pace, Vazquez and Richards	869968197049750	04/24	877	JCB 15 digit	andersonmichael@sherman.biz	Early years teacher	
9991	PSC 8165, Box 8498\nAPO AP 60327-0346	50 dA	AM	Mozilla/5.0 (compatible; MSIE 8.0; Windows NT ...	Snyder Inc	4221582137197481	02/24	969	Voyager	kking@wise-liu.com	IT sales professional	2
9992	885 Allen Mountains Apt. 230\nWallhaven, LA 16995	40 vH	PM	Mozilla/5.0 (Macintosh; PPC Mac OS X 10_6_5) A...	Wells Ltd	4664825258997302	10/20	431	Discover	bberry@wright.net	Set designer	1
9993	7555 Larson Locks Suite 229\nEllisburgh, MA 34...	72 jg	PM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_8...	Colon and Sons	30025560104631	10/25	629	Maestro	chelseawilliams@lopez.biz	Designer, exhibition/display	1
9994	6276 Rojas Hollow\nLake Louis, WY 56410-7837	93 Ex	PM	Opera/9.68.(X11; Linux x86_64; sl-SI) Presto/2...	Ritter-Smith	3112186784121077	01/25	1823	Maestro	iroberts@gmail.com	Education officer, museum	2
9995	966 Castaneda Locks\nWest Juliafurt, CO 96415	92 XI	PM	Mozilla/5.0 (Windows NT 5.1) AppleWebKit/5352 ...	Randall-Sloan	342945015358701	03/22	838	JCB 15 digit	iscott@wade-garner.com	Printmaker	2
9996	832 Curtis Dam Suite 785\nNorth Edwardburgh, T...	41 JY	AM	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Hale, Collins and Wilson	210033169205009	07/25	207	JCB 16 digit	mary85@hotmail.com	Energy engineer	12
9997	Unit 4434 Box 6343\nDPO AE 28026-0283	74 Zh	AM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_7...	Anderson Ltd	6011539787356311	05/21	1	VISA 16 digit	tyler16@gmail.com	Veterinary surgeon	1
9998	0096 English Rest\nRoystad, IA 12457	74 cL	PM	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8_8;...	Cook Inc	180003348082930	11/17	987	American Express	elizabethmoore@reid.net	Local government officer	
9999	40674 Barrett Stravenue\nGrimesville, WI 79682	64 Hr	AM	Mozilla/5.0 (X11; Linux i686; rv:1.9.5.20) Gec...	Greene Inc	4139972901927273	02/19	302	JCB 15 digit	rachelford@vaughn.com	Embryologist, clinical	176

Chapter 3 - Data Cleaning

Data cleaning is a crucial step in the data analysis process, as it ensures that the dataset is accurate, consistent, and ready for analysis.

- Handling missing values



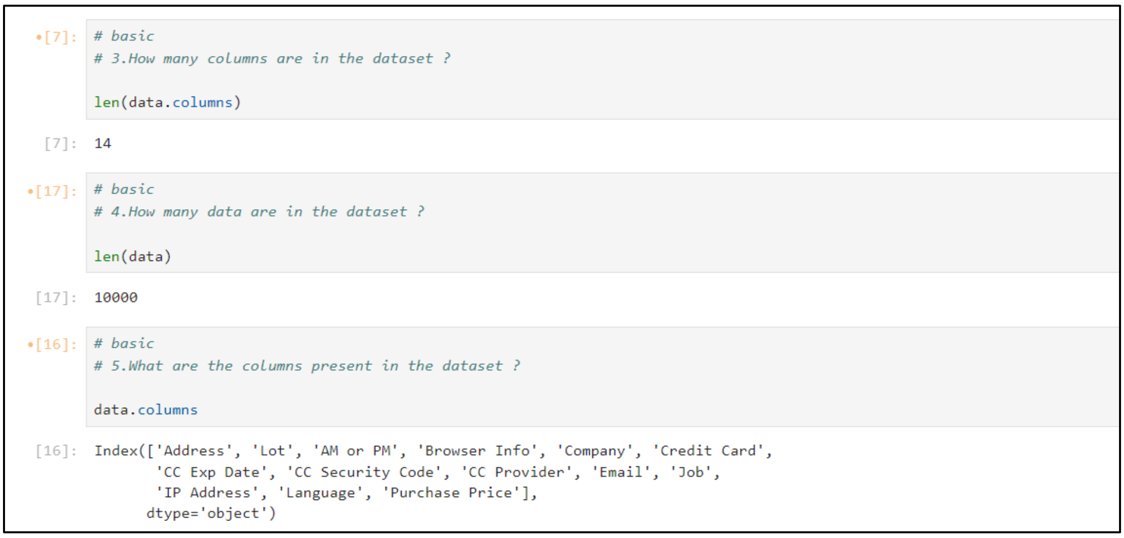
```
•[15]: # basic
       # 2.Check Null values in the dataset

       data.isnull().sum()

 [15]: Address             0
       Lot                 0
       AM or PM            0
       Browser Info        0
       Company             0
       Credit Card         0
       CC Exp Date         0
       CC Security Code    0
       CC Provider         0
       Email               0
       Job                 0
       IP Address          0
       Language            0
       Purchase Price      0
       dtype: int64
```

Here we verified whether any null values (missing values) present the “E-commerce Purchase” Dataset.

Chapter 4 - Exploratory Data Analysis (EDA)

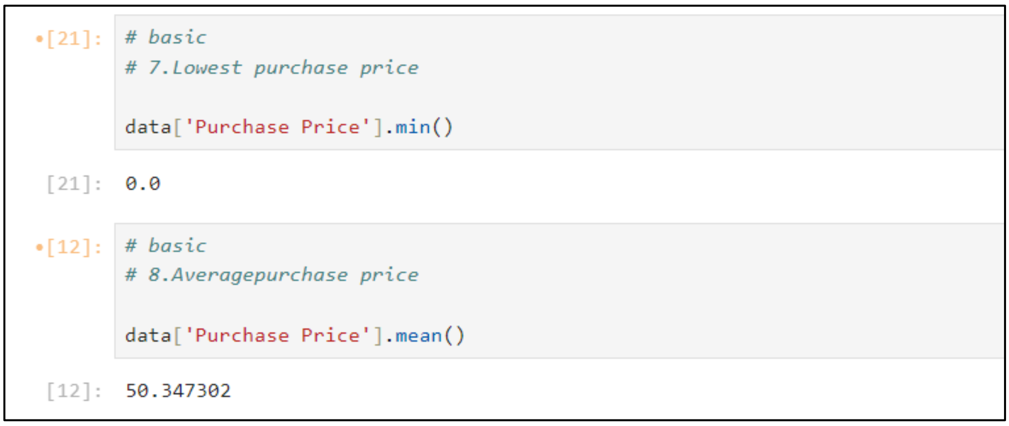
Exploratory Data Analysis (EDA) is an essential step in the data analysis process, as it helps us understand the structure of the dataset, identify patterns, and generate hypotheses for further analysis. In this project, we conducted a thorough EDA of the 'E-Commerce Purchase' dataset, using Pandas to perform basic and intermediate analyses.

I. Basic Analyses:



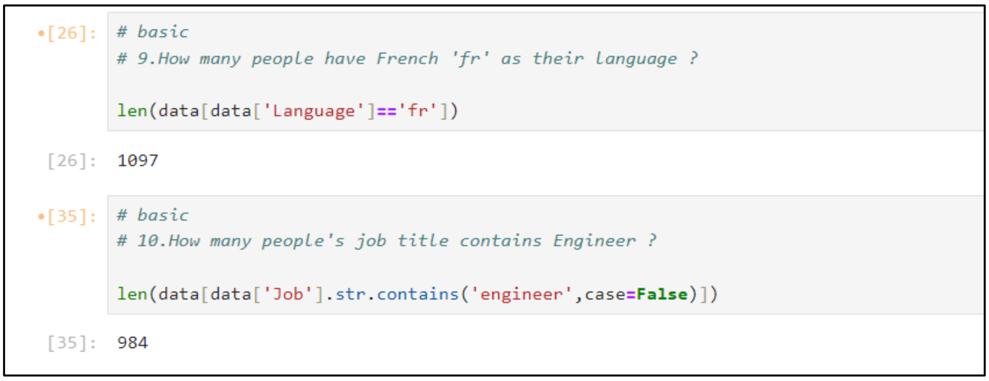
```
•[7]:  # basic
       # 3.How many columns are in the dataset ?

       len(data.columns)

 [7]:  14

•[17]: # basic
       # 4.How many data are in the dataset ?

       len(data)

 [17]: 10000

•[16]: # basic
       # 5.What are the columns present in the dataset ?

       data.columns

 [16]: Index(['Address', 'Lot', 'AM or PM', 'Browser Info', 'Company', 'Credit Card',
              'CC Exp Date', 'CC Security Code', 'CC Provider', 'Email', 'Job',
              'IP Address', 'Language', 'Purchase Price'],
             dtype='object')
```

- We explored the number of columns and rows in the dataset using the len() function.
- We identified the columns present in the dataset using the columns attribute.



```
•[21]:  # basic
        # 7.Lowest purchase price

        data['Purchase Price'].min()

 [21]:  0.0

•[12]:  # basic
        # 8.Averagepurchase price

        data['Purchase Price'].mean()

 [12]:  50.347302
```

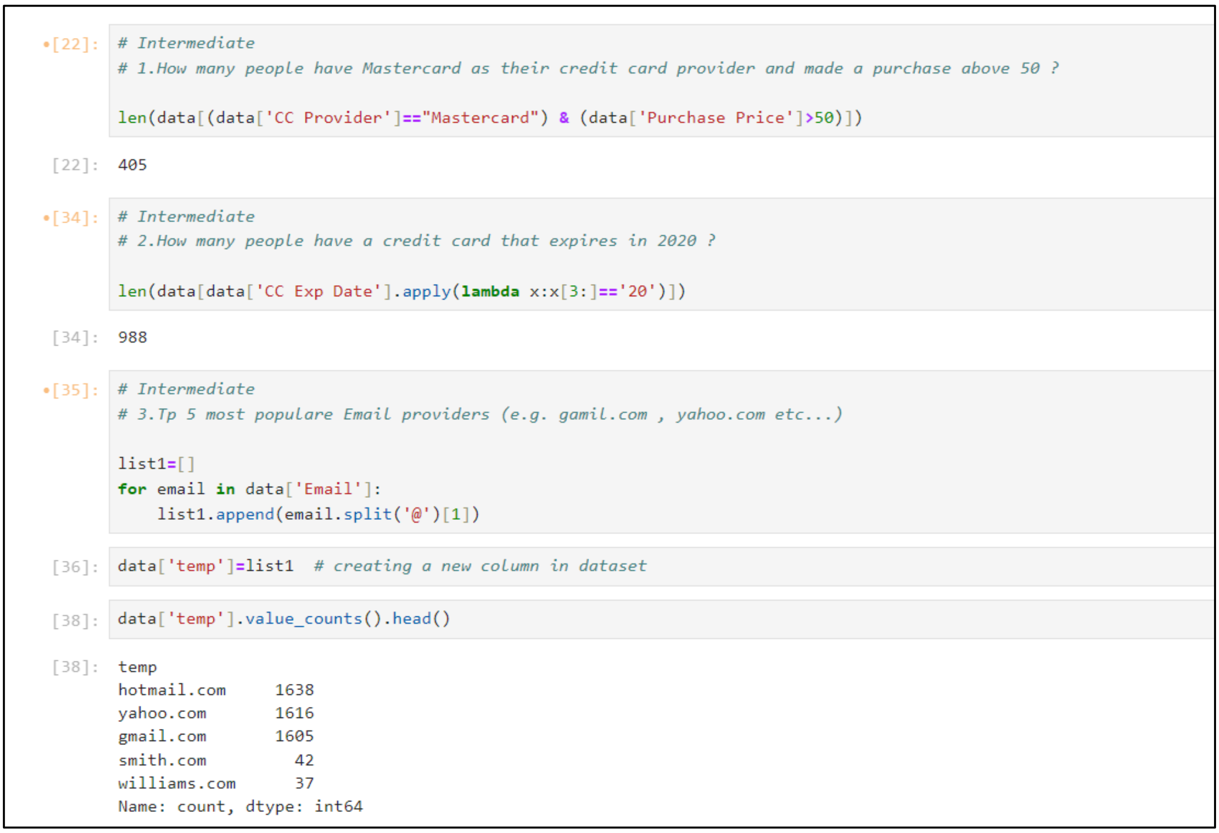
- We calculated summary statistics such as the highest, lowest purchase prices.



```
•[26]:  # basic
        # 9.How many people have French 'fr' as their language ?

        len(data[data['Language']=='fr'])

 [26]:  1097

•[35]:  # basic
        # 10.How many people's job title contains Engineer ?

        len(data[data['Job'].str.contains('engineer',case=False)])

 [35]:  984
```

- We counted the number of people with French as their language and job titles containing "Engineer."

II. Intermediate Analyses:



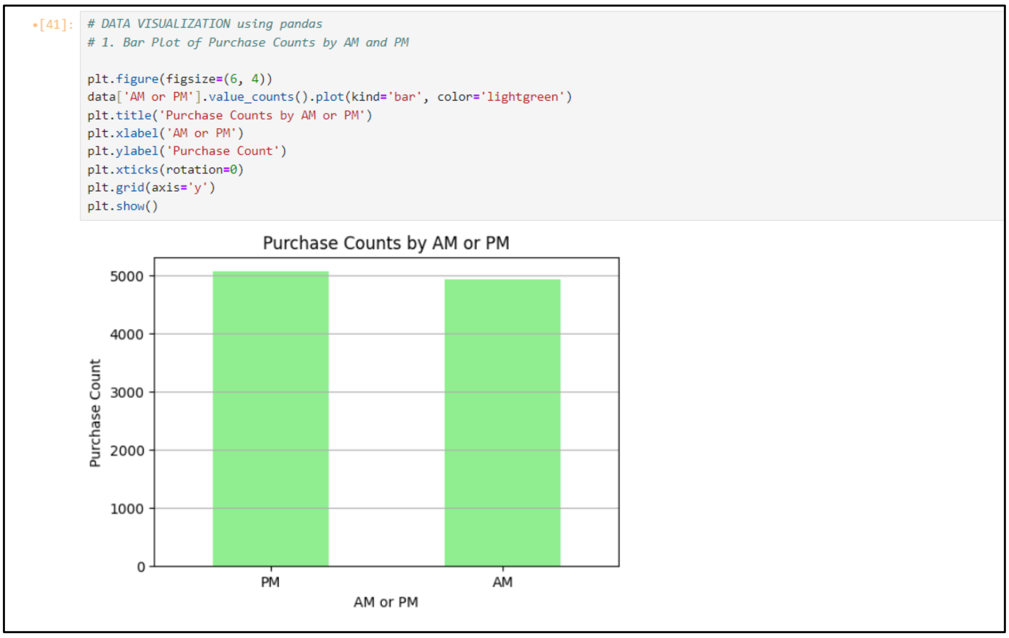
```
•[22]:  # Intermediate
        # 1.How many people have Mastercard as their credit card provider and made a purchase above 50 ?

        len(data[(data['CC Provider']=="Mastercard") & (data['Purchase Price']>50)])

 [22]:  405

•[34]:  # Intermediate
        # 2.How many people have a credit card that expires in 2020 ?

        len(data[data['CC Exp Date'].apply(lambda x:x[3:]=='20')])

 [34]:  988

•[35]:  # Intermediate
        # 3.Tp 5 most populare Email providers (e.g. gamil.com , yahoo.com etc...)

        list1=[]
        for email in data['Email']:
            list1.append(email.split('@')[1])

 [36]:  data['temp']=list1  # creating a new column in dataset

 [38]:  data['temp'].value_counts().head()

 [38]:  temp
        hotmail.com     1638
        yahoo.com       1616
        gmail.com       1605
        smith.com         42
        williams.com      37
        Name: count, dtype: int64
```

- We identified people with Mastercard as their credit card provider who made purchases above a certain threshold.
- We counted the number of people with a credit card that expires in 2020.
- We identified the top 5 most popular email providers.

Chapter 5 - Data Visualization

Data visualization plays a crucial role in understanding complex datasets, as it allows us to visually explore patterns, trends, and relationships that may not be apparent from the raw data alone. In this project, we used various data visualization techniques to enhance our understanding of the 'E-Commerce Purchase' dataset and communicate our findings effectively.

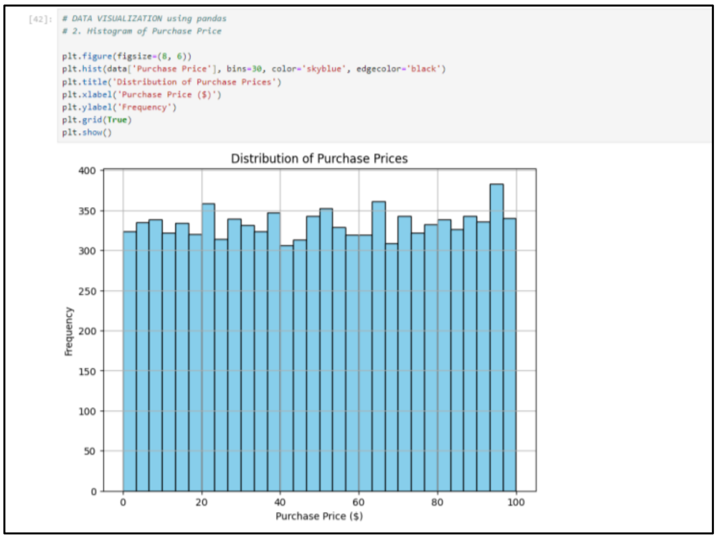
I. Bar Plot of Purchase Counts by AM and PM:



```
•[41]: # DATA VISUALIZATION using pandas
       # 1. Bar Plot of Purchase Counts by AM and PM

       plt.figure(figsize=(6, 4))
       data['AM or PM'].value_counts().plot(kind='bar', color='lightgreen')
       plt.title('Purchase Counts by AM or PM')
       plt.xlabel('AM or PM')
       plt.ylabel('Purchase Count')
       plt.xticks(rotation=0)
       plt.grid(axis='y')
       plt.show()
```

- This visualization helps us understand the distribution of purchases made during the day and night.
- It contributes to the analysis by highlighting any trends in purchasing behavior based on the time of day.

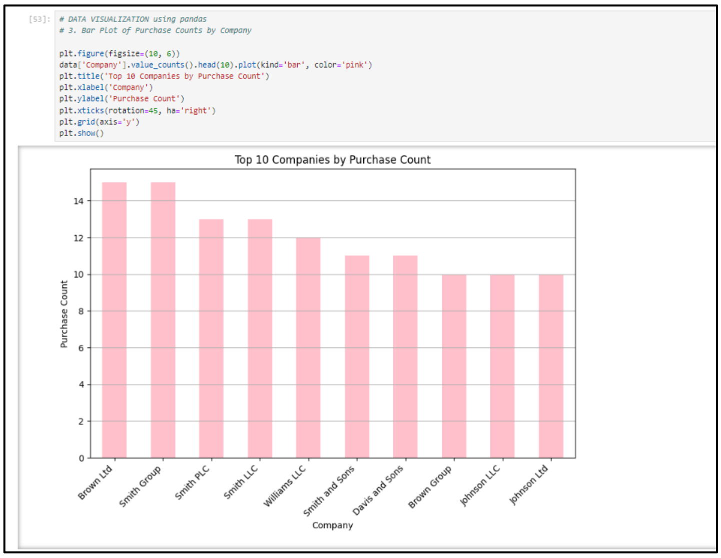
II. Histogram of Purchase Price:



```
# DATA VISUALIZATION using pandas
# 2. Histogram of Purchase Price

plt.figure(figsize=(8, 6))
plt.hist(data['Purchase Price'], bins=30, color='skyblue', edgecolor='black')
plt.title('Distribution of Purchase Prices')
plt.xlabel('Purchase Price ($)')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```

- The histogram provides a visual representation of the distribution of purchase prices.
- It helps us identify the most common purchase price range and any outliers in the data.

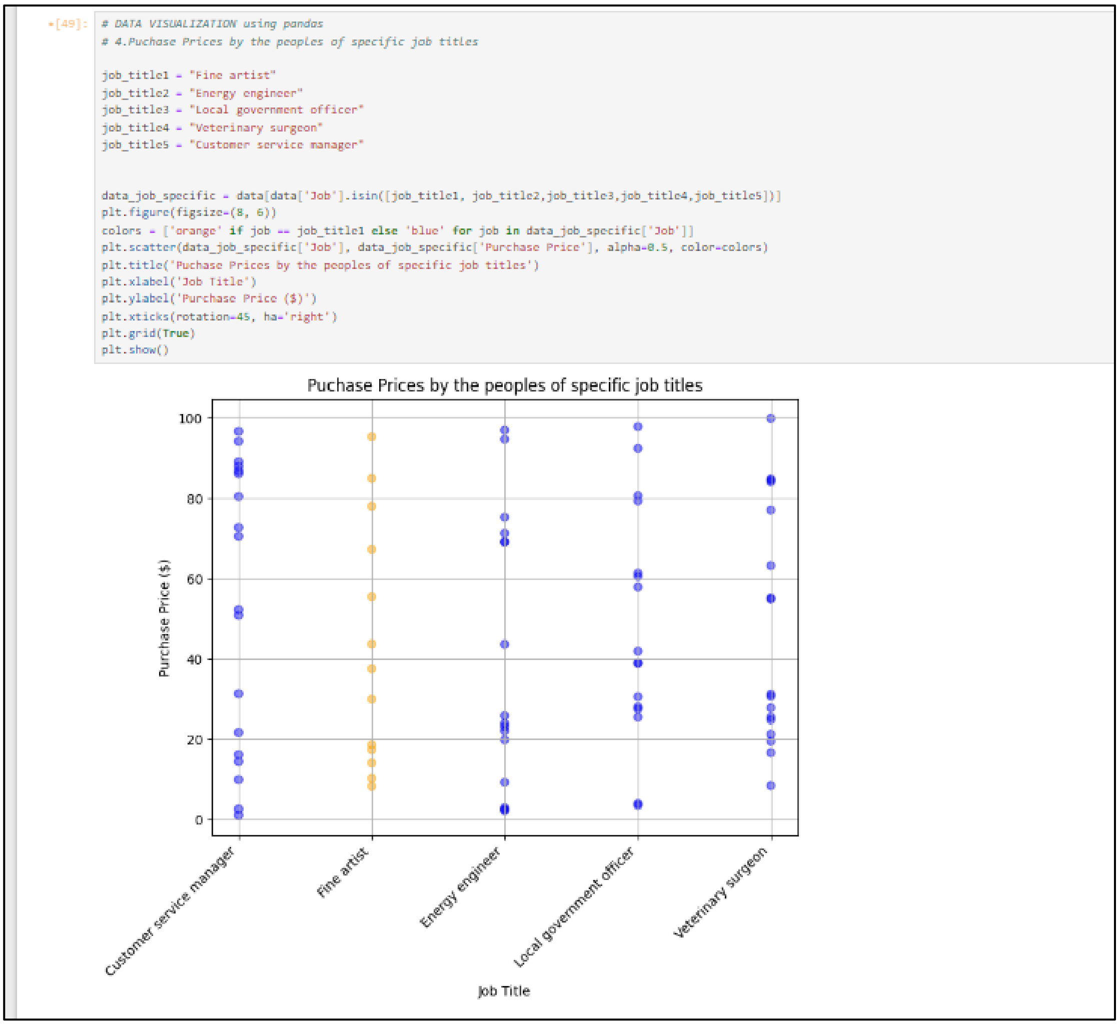
III. Bar Plot of Purchase Counts by Company:



```
# DATA VISUALIZATION using pandas
# 3. Bar Plot of Purchase Counts by Company

plt.figure(figsize=(10, 6))
data['Company'].value_counts().head(10).plot(kind='bar', color='pink')
plt.title('Top 10 Companies by Purchase Count')
plt.xlabel('Company')
plt.ylabel('Purchase Count')
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y')
plt.show()
```

- This visualization allows us to compare the purchase counts across different companies.
- It contributes to the analysis by highlighting the popularity of certain companies among customers.

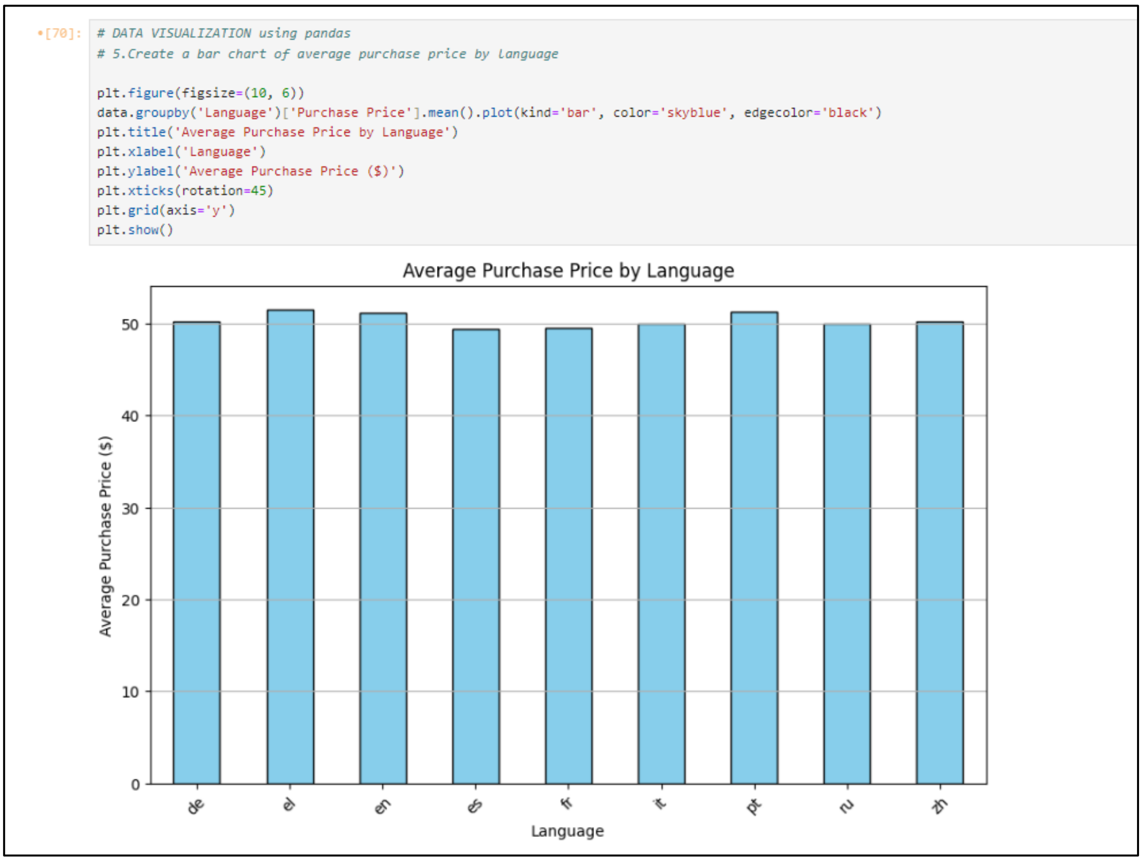
IV. Scatter Plot of Purchase Prices by Specific Job Titles:



```
•[49]: # DATA VISUALIZATION using pandas
       # 4.Puchase Prices by the peoples of specific job titles

       job_title1 = "Fine artist"
       job_title2 = "Energy engineer"
       job_title3 = "Local government officer"
       job_title4 = "Veterinary surgeon"
       job_title5 = "Customer service manager"


       data_job_specific = data[data['Job'].isin([job_title1, job_title2,job_title3,job_title4,job_title5])]
       plt.figure(figsize=(8, 6))
       colors = ['orange' if job == job_title1 else 'blue' for job in data_job_specific['Job']]
       plt.scatter(data_job_specific['Job'], data_job_specific['Purchase Price'], alpha=0.5, color=colors)
       plt.title('Puchase Prices by the peoples of specific job titles')
       plt.xlabel('Job Title')
       plt.ylabel('Purchase Price ($)')
       plt.xticks(rotation=45, ha='right')
       plt.grid(True)
       plt.show()
```

- The scatter plot helps us visualize the relationship between purchase prices and specific job titles.
- It contributes to the analysis by identifying any patterns or trends in purchasing behavior based on occupation.

V. Bar Chart of Average Purchase Price by Language:

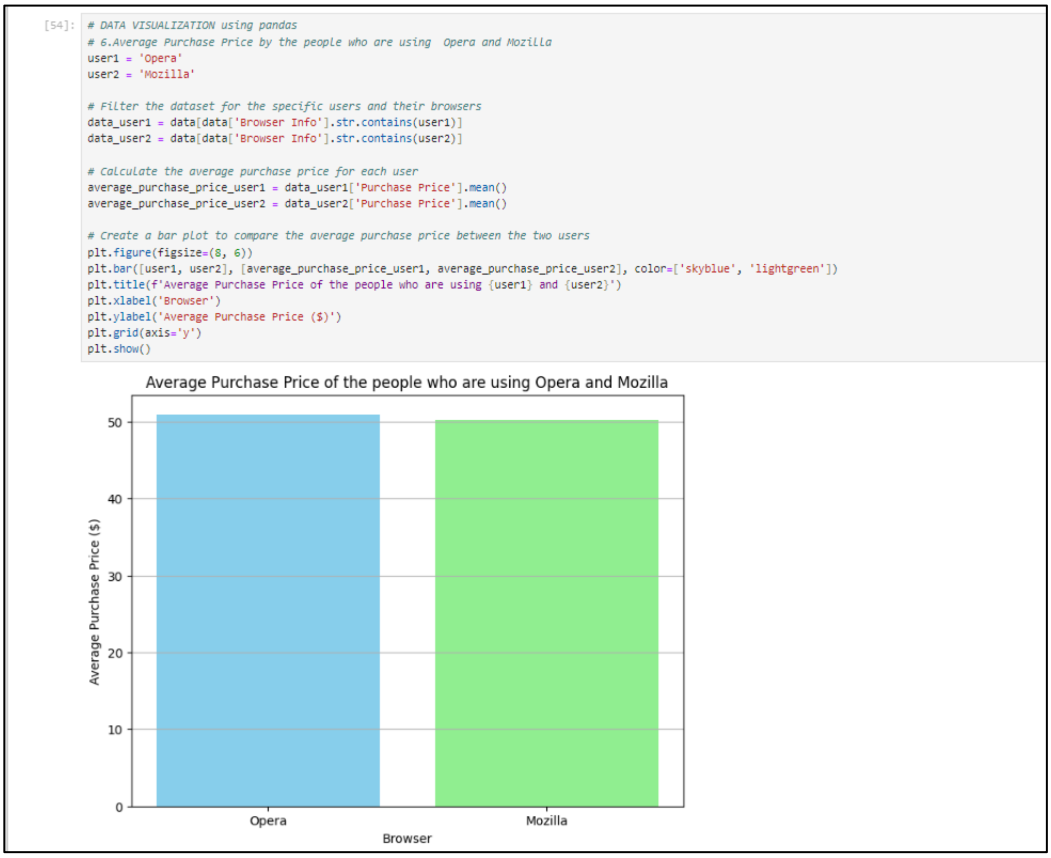


```
•[70]: # DATA VISUALIZATION using pandas
       # 5.Create a bar chart of average purchase price by language

       plt.figure(figsize=(10, 6))
       data.groupby('Language')['Purchase Price'].mean().plot(kind='bar', color='skyblue', edgecolor='black')
       plt.title('Average Purchase Price by Language')
       plt.xlabel('Language')
       plt.ylabel('Average Purchase Price ($)')
       plt.xticks(rotation=45)
       plt.grid(axis='y')
       plt.show()
```

Average Purchase Price by Language
Average Purchase Price ($)
0
10
20
30
40
50
de
el
en
es
fr
it
pt
ru
zh
Language

- This visualization shows the average purchase price for each language.
- It contributes to the analysis by highlighting any differences in spending habits among customers speaking different languages.

VI. Bar Chart of Average Purchase Price by Browser:



```
[54]: # DATA VISUALIZATION using pandas
      # 6.Average Purchase Price by the people who are using  Opera and Mozilla
      user1 = 'Opera'
      user2 = 'Mozilla'

      # Filter the dataset for the specific users and their browsers
      data_user1 = data[data['Browser Info'].str.contains(user1)]
      data_user2 = data[data['Browser Info'].str.contains(user2)]

      # Calculate the average purchase price for each user
      average_purchase_price_user1 = data_user1['Purchase Price'].mean()
      average_purchase_price_user2 = data_user2['Purchase Price'].mean()

      # Create a bar plot to compare the average purchase price between the two users
      plt.figure(figsize=(8, 6))
      plt.bar([user1, user2], [average_purchase_price_user1, average_purchase_price_user2], color=['skyblue', 'lightgreen'])
      plt.title(f'Average Purchase Price of the people who are using {user1} and {user2}')
      plt.xlabel('Browser')
      plt.ylabel('Average Purchase Price ($)')
      plt.grid(axis='y')
      plt.show()
```

Average Purchase Price of the people who are using Opera and Mozilla
Average Purchase Price ($)
0
10
20
30
40
50
Opera
Mozilla
Browser

- This visualization compares the average purchase price among customers using different browsers.
- It contributes to the analysis by identifying any differences in purchasing behavior based on the browser used.

Chapter 6 - Key Findings

After conducting a comprehensive analysis of the 'E-Commerce Purchase' dataset, several key findings emerged, shedding light on customer behavior and trends within the e-commerce platform:

Purchase Patterns:

- The majority of purchases occurred during the day (AM) rather than at night (PM), indicating a potential trend in shopping behavior.
- The distribution of purchase prices was right-skewed, with most purchases falling within a lower price range.

Credit Card Usage:

- Customers using Mastercard as their credit card provider tended to make purchases above a certain threshold, indicating potential differences in spending habits based on payment method.
- A significant number of customers had credit cards expiring in 2020, suggesting the need for targeted marketing campaigns to retain these customers.

Company Engagement:

- Purchase counts varied widely among companies, with some companies having significantly higher purchase counts than others, indicating varying levels of engagement with the platform.

Language Preferences:

- The average purchase price varied among customers speaking different languages, suggesting differences in spending habits based on language.

Browser Usage:

- Customers using different browsers exhibited differences in average purchase price, highlighting the potential impact of browser choice on purchasing behavior.

Overall, these findings provide valuable insights into customer behavior and preferences within the e-commerce platform, offering opportunities for targeted marketing strategies, product offerings, and customer engagement initiatives.

Conclusion

- The analysis of the 'E-Commerce Purchase' dataset using Pandas has provided valuable insights into customer behavior and trends within then e-commerce platform. Through data cleaning, exploratory data analysis (EDA), and data visualization, we were able to uncover key patterns and trends that can inform business decisions and strategies.

- Overall, these findings can help e-commerce businesses better understand their customers and tailor their offerings to meet their needs, ultimately leading to improved customer satisfaction and business success.

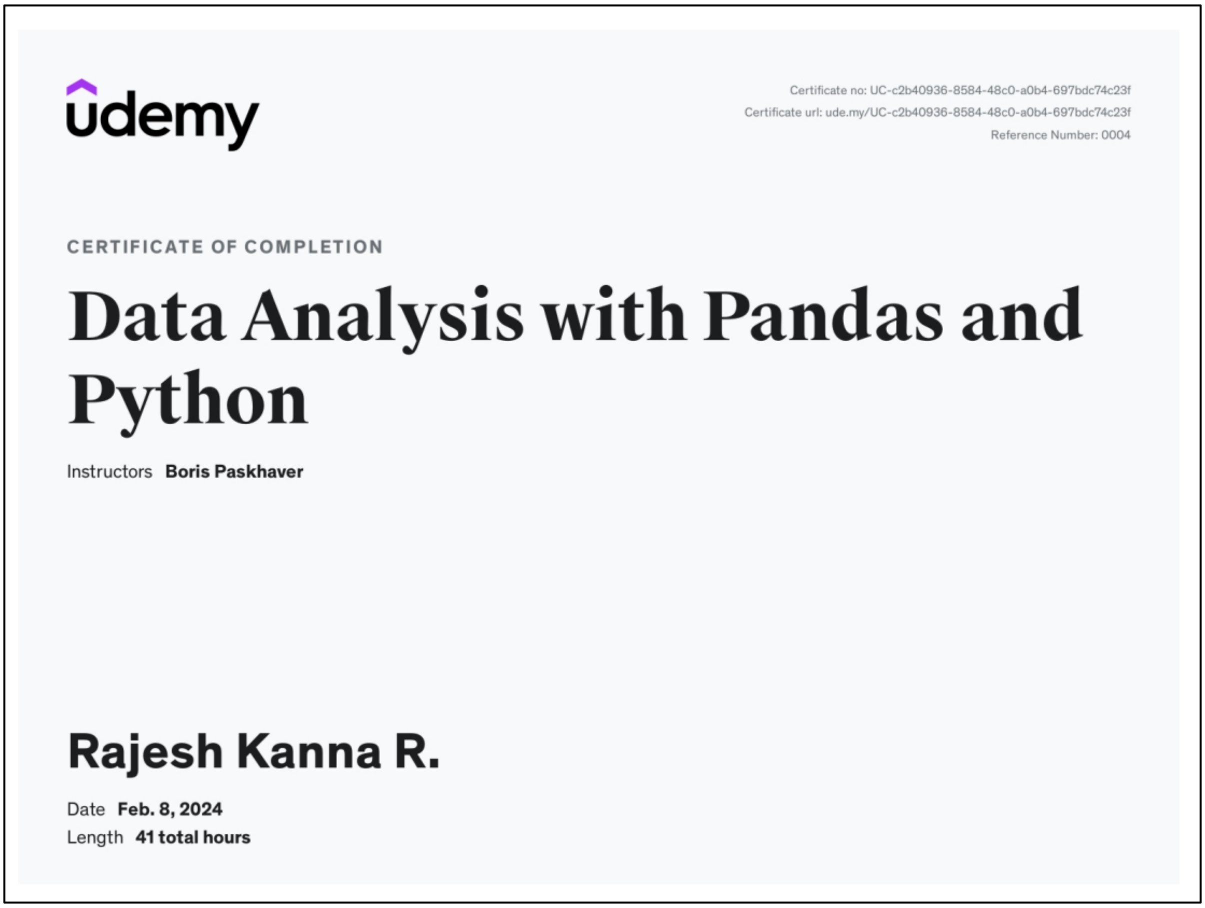
- In conclusion, the analysis of the 'E-Commerce Purchase' dataset has demonstrated the power of data analysis and visualization in extracting meaningful insights from complex datasets. By leveraging these insights, businesses can make informed decisions that drive growth and enhance the customer experience.

Reference

Dataset link:

https://www.kaggle.com/datasets/utkarsharya/ecommerce-purchases

Certification



udemy

Certificate no: UC-c2b40936-8584-48c0-a0b4-697bdc74c23f
Certificate url: ude.my/UC-c2b40936-8584-48c0-a0b4-697bdc74c23f
Reference Number: 0004

CERTIFICATE OF COMPLETION

Data Analysis with Pandas and Python

Instructors **Boris Paskhaver**

Rajesh Kanna R.

Date **Feb. 8, 2024**
Length **41 total hours**