ÇUKUROVA UNIVERSITY

ENGINEERING AND ARCHITECTURE FACULTY

COMPUTER ENGINEERING

GRADUATION THESIS

OPEN BANK TRANSACTION DATA VISUALIZATION

By

2019556043/İbrahim Doğu KOCAÇİFTÇİ

Advisor

Öğr. Gör. Dr. Elif Emel FIRAT

Ay-Yıl

June 2023

ADANA

Open Bank Transaction Data Visualization

Abstract:

This thesis delves into the exploration and visualization of Open Bank Transaction Data, an extensive dataset encompassing an individual's bank transactions spanning a period of seven years, from 2015 to 2022. The primary objective of this research is to leverage data visualization techniques to enhance the comprehension, analysis, and interpretation of this wealth of financial data.

Data visualization serves as a fundamental tool in converting intricate numerical data into visual representations that are easily accessible and meaningful. By harnessing the potential of the Open Bank Transaction Data, this study undertakes a comprehensive investigation into the application of diverse visualization methods, with the aim of uncovering patterns, trends, and insightful observations pertaining to individual financial behaviors.

Among the various visualization techniques employed, treemaps emerge as a prominent and powerful tool. Treemaps offer a hierarchical depiction of transaction data, facilitating the intuitive exploration of financial patterns at varying levels of detail. Through the utilization of treemaps, it becomes possible to discern spending categories, track transaction volumes, and identify temporal trends. Furthermore, treemaps provide a comprehensive overview of financial activities, enabling the comparison of spending patterns across different time periods and categories.

The Open Bank Transaction Data, being a longitudinal dataset spanning several years, presents a unique opportunity to examine the evolution of individual financial behaviors and gain valuable insights into saving habits, spending patterns, and investment trends. By employing data visualization techniques, this research endeavors to uncover significant findings, including seasonal variations in spending behavior, shifts in expenditure patterns, and potential correlations between financial activities and external factors.

Through the thorough analysis of the Open Bank Transaction Data using visualization techniques, this thesis contributes to the fields of computer engineering and financial analysis by providing a detailed examination of individual financial behaviors over an extended timeframe. The findings derived from this research hold implications for personal finance management, financial planning, and decision-making processes, ultimately empowering individuals to make well-informed choices regarding their financial well-being. By shedding light on the intricacies of individual financial behaviors, this research paves the way for enhanced financial literacy, facilitating improved financial decision-making in an ever-evolving financial landscape.

Contents

Figure List

1. INTRODUCTION:

In today's data-driven world, the ability to effectively visualize and interpret complex datasets is paramount. Data visualization serves as a powerful tool for gaining insights, identifying patterns, and making informed decisions. In this thesis, our focus lies in the realm of data visualization, specifically in the context of Open Bank Transaction Data. We aim to employ treemaps, a visually compelling technique, to create informative and analytical representations of financial activities. Through the visualization of five distinct stories derived from the dataset, we seek to shed light on various aspects of financial behaviors and patterns.

The motivation behind this research stems from multiple sources. Firstly, the unwavering trust and guidance of our thesis advisor have encouraged us to delve into this captivating area of study. Their belief in our abilities has instilled a sense of purpose and determination to undertake this research endeavor. Moreover, the intrinsic fascination of exploring the intricacies of financial transactions and uncovering hidden insights has propelled our interest in this subject matter. Additionally, a personal drive to expand our knowledge and skills in the field of data visualization has been a driving force behind this research. We recognize the immense potential that data visualization holds in enhancing financial literacy, aiding financial planning, and facilitating decision-making processes. By effectively visualizing and analyzing the Open Bank Transaction Data, we aim to contribute to these domains and empower individuals to make informed financial choices.

The scope of our research is centered around the accurate and analytical visualization of the selected data stories. Our exploration revolves around five key narratives derived from the dataset: a comparative analysis between BP (Bill Payments) and other transaction types spanning the years 2018-2022, an examination of user transaction activity, an investigation into the relationship between balance fluctuations and transaction types across various locations on a monthly basis, an exploration of major expense categories categorized by transaction types and location, and a comparison between quarantine and non-closed periods from December 1, 2019, to December 31, 2022.

Fortuitously, we have had the privilege of working with a meticulously maintained dataset that offers comprehensive and reliable information. We have not encountered any significant limitations, such as missing or incomplete details, during our analysis. The regularity of the data collection process has further bolstered the robustness and reliability of our findings.

In terms of data visualization techniques and analytical approaches, there are no restrictions or specific considerations. This flexibility has allowed us to explore various visualization methods and employ advanced analytical techniques to derive meaningful insights from the data. However, the selection of the subjects to visualize posed a challenge. With a plethora of potential stories to tell, it was essential to carefully curate the most informative and impactful narratives. By choosing these subjects judiciously, we aim to present a comprehensive understanding of individual financial behaviors and patterns, enabling individuals to gain valuable insights into their financial activities.

The subsequent sections of this thesis are structured in a logical and coherent manner. Chapter 2 provides an extensive literature review, delving into relevant studies, research papers, and scholarly articles in the field of data visualization, financial analysis, and related techniques. This comprehensive exploration aims to establish the foundation for our research, identify existing knowledge gaps, and highlight potential research opportunities. In Chapter 3, we delineate the methodology employed in this study. This encompasses data preprocessing techniques, analytical approaches, and the implementation of treemaps for visualization. Chapter 4 presents the five data stories we have chosen, accompanied by the corresponding treemap visualizations. These visualizations serve as a medium through which the intricate patterns and relationships within the data can be effectively conveyed and understood. In Chapter 5, we embark on a detailed analysis of the insights derived from the treemap visualizations, providing valuable interpretations and implications for financial literacy and decision-making. Finally, Chapter 6 encapsulates the thesis by summarizing the key findings, discussing the contributions made by this research, and suggesting potential avenues for future exploration and development.

This research holds great significance in the context of financial literacy and decision-making. The visualization of Open Bank Transaction Data not only provides a comprehensive overview of financial activities but also enables individuals to assess their financial health, identify trends, and make informed choices. By leveraging the power of treemaps, we can present complex financial information in a visually appealing and easily interpretable manner, fostering better comprehension and engagement.

Moreover, the utilization of treemaps as a visualization technique offers several advantages. Treemaps provide a hierarchical representation of data, allowing users to navigate through different levels of information and gain insights into the relationships between various dimensions. The use of color coding and size proportions in treemaps enables the effective communication of data patterns and distributions. Additionally, treemaps facilitate the comparison of different categories, highlighting variations and anomalies within the dataset. These attributes make treemaps a valuable tool for visualizing and analyzing financial data, enabling users to grasp complex patterns and trends effortlessly.

However, it is important to acknowledge the limitations of this research. The findings and interpretations drawn from the Open Bank Transaction Data are contingent upon the quality and accuracy of the dataset itself. While efforts have been made to ensure data integrity and protection, it is essential to recognize that data collected from various sources may contain inherent biases, errors, or omissions. The generalizability of the findings may also be influenced by the specific characteristics of the dataset used. Therefore, caution should be exercised when extrapolating the results to broader populations or contexts.

Validity and reliability are paramount considerations in any research endeavor. To address these aspects, stringent measures have been implemented throughout the research process. Data cleansing techniques have been employed to minimize errors and inconsistencies in the dataset. Robust analytical methods, guided by established statistical principles, have been utilized to ensure the reliability and accuracy of the findings. Additionally, the utilization of treemaps as a visualization technique aligns with established best practices and standards in the field of data visualization.

In conclusion, this thesis embarks on a journey to explore and visualize the Open Bank Transaction Data through the use of treemaps. By providing comprehensive insights into financial activities, patterns, and relationships, this research contributes to the fields of data visualization and financial analysis. The utilization of treemaps as a visualization technique enables individuals to gain a deeper understanding of their financial behaviors, enhancing financial literacy and supporting informed decision-making. Through careful methodology, robust analysis, and thoughtful interpretation, this research strives to unlock the stories hidden within the dataset, empowering individuals to navigate their financial landscapes with confidence and knowledge.

2. LIERATURE REVIEW:

The Literature Review section provides an overview of the existing knowledge related to data visualization and financial analysis. It highlights the significance of data visualization in finance and the importance of analyzing financial data to make informed decisions.

Data Visualization:

Effective data visualization techniques enable individuals to understand complex financial information. Visual representations like treemaps, bar charts, and line graphs are commonly used to present financial data in a clear and intuitive manner.

Financial Data Analysis:

Analyzing financial data helps uncover patterns and relationships in individual financial behaviors. Techniques such as exploratory data analysis, clustering, and regression analysis are employed to gain insights into financial trends and assist in decision-making processes.

Related Studies:

Previous research has focused on visualizing financial market data, analyzing personal spending behaviors, and developing forecasting models for financial time series data. These studies emphasize the importance of interactive visualizations, data preprocessing, and user-centered design approaches.

Synthesis and Analysis:

The reviewed literature suggests that data visualization and financial analysis can improve financial literacy and decision making. However, there is a need for further research on visualizing Open Bank Transaction Data specifically. The integration of real-time data and dynamic visualizations also poses challenges that require exploration.

Conclusion:

The Literature Review section presents an overview of data visualization and financial analysis. It establishes the importance of these fields in finance and identifies gaps in the literature regarding the visualization of Open Bank Transaction Data. In the subsequent sections, we will utilize treemaps to visualize the selected data stories, contributing to the field of data visualization and financial analysis.

2.1. Methodology:

The Methodology section plays a pivotal role in elucidating the research design and methodology employed throughout this study. It provides a detailed account of the data collection methods, tools, and techniques utilized, justifies the chosen approach, and delves into the limitations, validity, and reliability of the research methods.

Research Design:

In order to comprehensively examine the Open Bank Transaction Data, a mixed-method research design was thoughtfully selected. While quantitative analysis took precedence, a qualitative perspective was also strategically incorporated to extract nuanced insights and pose relevant questions. This synergistic combination allowed for a holistic analysis and interpretation of the dataset, resulting in a comprehensive understanding of the phenomena under investigation.

Data Collection Methods:

The Open Bank Transaction Data used in this study was meticulously acquired from a publicly available research dataset. Emphasis was placed on ensuring data integrity and protection, adhering to strict ethical guidelines and complying with rigorous data usage policies. Anonymization and aggregation techniques were meticulously applied to safeguard the privacy and confidentiality of individuals involved in the transactions, thereby maintaining the utmost level of data security.

Tools and Techniques:

For the purpose of data analysis and manipulation, the highly versatile and robust pandas library in Python was judiciously employed. This powerful library enabled a systematic and thorough exploration of the dataset, facilitating efficient data processing, cleaning, and transformation. Additionally, RawGraphs 2.0, an intuitive web-based tool, proved instrumental in gaining preliminary insights and generating treemaps, which offered a visually compelling and insightful representation of the data. Furthermore, to enhance the visualization aspect of the study, the widely acclaimed Plotly library was utilized, enabling the creation of interactive and visually captivating charts. The utilization of Python as the programming language ensured flexibility and ease of integration with other data analysis tools and techniques.

Justification of Approach:

The chosen methodology harmoniously aligns with the research objectives, providing a robust framework for an in-depth analysis of the Open Bank Transaction Data. The mixed-method approach, seamlessly integrating quantitative and qualitative analysis, presented numerous advantages. Quantitative analysis enabled statistical examination, unveiling trends, patterns, and correlations within the dataset. On the other hand, qualitative analysis fostered a deeper understanding and interpretation of the data, capturing the intricacies and nuances that quantitative measures alone may overlook. This duality of approach enhanced the richness, reliability, and validity of the research findings, reinforcing the overall robustness of the study.

Limitations:

It is imperative to acknowledge and address the limitations inherent in this study. Firstly, the reliance on the provided Open Bank Transaction Data introduces potential biases or limitations specific to the dataset itself. It is crucial to approach the findings with the understanding that they may not fully encapsulate the entirety of the population or reflect broader trends beyond the dataset's scope. Secondly, the presence of missing or incomplete information within the dataset may impact the analysis and interpretation of the findings, necessitating caution in drawing definitive conclusions. Moreover, the study is confined to the specific time frame of 2015-2022, which poses inherent temporal limitations. The findings may not account for potential changes or shifts in the financial landscape beyond the designated period.

Validity and Reliability:

To ensure the validity of the research findings, stringent measures were implemented throughout the data analysis process. Robust data analysis techniques, adhering to established methodologies, were meticulously employed to enhance the internal validity of the study. Furthermore, the utilization of widely recognized tools and libraries, such as pandas and Plotly, contributed to the reliability of the outcomes. Additionally, inter-rater reliability checks were performed to validate the consistency of the interpretations and findings, further bolstering the validity and reliability of the research methods. It is important to recognize that while efforts were made to enhance the validity and reliability of the study, inherent limitations and constraints exist within the dataset and the research context.

Conclusion:

The Methodology section provided an in-depth exploration of the research design, data collection methods, and tools employed in this study. The utilization of a mixed-method approach, incorporating both quantitative and qualitative analysis, facilitated a comprehensive examination of the Open Bank Transaction Data. The implementation of the pandas library, RawGraphs 2.0, and Plotly enabled efficient data manipulation, insightful visualizations, and interactive charting. Despite acknowledging the limitations, rigorous measures were adopted to uphold the validity and reliability of the research methods. By leveraging a thoughtfully chosen methodology, this study lays a solid foundation for the subsequent data analysis and exploration in the pursuit of meaningful insights from the Open Bank Transaction Data.

2.2. Results and Analysis:

In this section, we present the findings obtained from the analysis of the Open Bank Transaction Data. The results are organized and presented in a clear and concise manner, using graphs, charts, and tables to illustrate the key insights derived from the dataset.
The dataset used for this analysis consists of an Excel file containing ten distinct columns: Transaction Number, Transaction Date, Transaction Type, Transaction Description, Debit Amount, Credit Amount, Category, Location City, and Location Country. Each column represents specific attributes and values related to the bank transactions.
Of particular importance for this analysis is the Transaction Type column, which contains abbreviations representing different transaction types within the dataset. To facilitate our analysis, it is crucial to familiarize ourselves with these abbreviations to understand the specific transaction types. The following abbreviations and their corresponding meanings are relevant to our analysis:

BGC: Bank Giro Credit
BP: Bill Payment
C/P: Cash Point
CHQ: Cheque
D/D: Direct Debit
DEB: Payment type Debit Card
FEE: Fixed Service Charge
FPI: Faster Payments Inwards
FPO: Faster Payments Outwards
PAY: Payment
TFR: Transfer
SO: Standing Order

2.3. Comparison Of Bp (Invoice Payments) And Other Transaction Types (2018-2022):

To compare BP (Bill Payments) with other transaction types, a treemap visualization was created using Python, Pandas, and Plotly. The following steps were followed to obtain the visualization:

```python
import pandas as pd

def xlsx_to_csv(xlsx_file, csv_file):

    df = pd.read_excel(xlsx_file)

    df.to_csv(csv_file, index=False)

xlsx_file = 'tez.xlsx'
csv_file = 'your_dataset.csv'

xlsx_to_csv(xlsx_file, csv_file)
```

Fig. 2.1. (Python Script That Converts xlsx to csv)

```python
import pandas as pd
import plotly.express as px

# Read the dataset
df = pd.read_csv('your_dataset.csv')

# Convert the date column to datetime type
df['Transaction Date'] = pd.to_datetime(df['Transaction Date'], dayfirst=True)

# Extract the year and month from the transaction date
df['YearMonth'] = df['Transaction Date'].dt.to_period('M')

# Filter data for the years 2018-2022
df_filtered = df[df['Transaction Date'].dt.year.between(2018, 2022)]

# Group data by location country, location city, year and month, and transaction type
grouped_data = df_filtered.groupby(['Location Country', 'Location City', 'YearMonth', 'Transaction Type']).size().reset_index(name='Transaction Count')

# Filter data for "BP" transaction type
bp_data = grouped_data[grouped_data['Transaction Type'] == 'BP']

# Create the treemap
fig = px.treemap(grouped_data,
                 path=['Location Country', 'Location City', 'YearMonth', 'Transaction Type'],
                 values='Transaction Count',
                 color='Transaction Count',
                 color_continuous_scale='viridis',
                 title='Comparison of BP with Other Transaction Types (2018-2022)')

# Set hover template
fig.update_traces(hovertemplate=
    '<b>Location Country:</b> %{label[0]}<br>' +
    '<b>Location City:</b> %{label[1]}<br>' +
    '<b>Year-Month:</b> %{label[2]}<br>' +
    '<b>Transaction Type:</b> %{label[3]}<br>' +
    '<b>Transaction Count:</b> %{value}<extra></extra>')

# Set layout options
fig.update_layout(height=1080, width=1920)

# Show the treemap
fig.show()
```

Fig. 2.2. (The Main Code)

The dataset was read, and the necessary libraries were imported.

The transaction date column was converted to the datetime type to facilitate further analysis. This step was crucial to focus the visualization on a monthly basis, allowing for a clearer comparison.

The year and month were extracted from the transaction date, providing a temporal dimension to the analysis.

The data were filtered for the years 2018-2022, ensuring a specific timeframe for the comparison and creating a more comprehensive visual representation.

The data were grouped by location country, location city, year and month, and transaction type. This grouping facilitated the hierarchical structure of the treemap, enabling a detailed comparison.

The data were further filtered to isolate the "BP" (Bill Payment) transaction type for specific analysis.

The treemap was created based on the defined hierarchy, illustrating the distribution and proportions of transaction types across different locations and time periods.
The hover template was set to enhance the understanding of the data. By hovering over specific regions of the treemap, relevant information and data details were displayed, allowing for in-depth exploration and analysis.
Layout options were adjusted to optimize the visual representation, and the treemap was displayed.

Upon analyzing the treemap visualization, several key insights were derived:

Fig. 2.3. (The Treemap)

Geographic Distribution: The hierarchical structure of the treemap, with location country and city as primary categories, revealed that the individual predominantly conducted transactions within the country without venturing abroad. Notably, Nottingham City had a higher transaction count compared to Swansea City, suggesting that the person's activities were more concentrated in Nottingham.

Temporal Analysis: Analyzing the date ranges provided valuable insights. The majority of transactions in Swansea City occurred in 2020 and earlier, indicating a potential shift in residence to Nottingham. This observation implies that the person may have settled in Nottingham and conducted more financial activities there.

Cost of Living: The comparison between Nottingham and Swansea could provide insights into the cost of living. Since Nottingham had a larger transaction count, it can be inferred that the individual incurred more expenses in Nottingham. This suggests that Nottingham may offer a more expensive city life compared to Swansea.

Transaction Types: Examining the transaction types within the hierarchy, it was observed that the person primarily utilized the DEB (Payment type Debit Card) category, indicating a preference for spending money through debit card transactions. Additionally, the treemap highlighted that BP (Bill Payment) occupied a significant portion of the transaction count, second only to the DEB type each month. This finding suggests that BP transactions held significant importance for the individual, representing a considerable portion of their financial activities.

By visualizing and analyzing the dataset, valuable insights were gained regarding the comparison of BP (Invoice Payments) with other transaction types. The treemap provided a comprehensive overview of transaction patterns, enabling an in-depth understanding of financial behaviors and preferences. These insights contribute to enhancing financial literacy and decision-making processes, empowering individuals to make informed choices regarding their financial activities.

2.4. User’s Transaction Activity:

In this section, we delve into the user's transaction activity to analyze patterns and differences based on the days of the week. By examining transaction frequency, volume, and distribution across different categories, we aim to identify notable trends or anomalies in the user's financial behavior.

```python
import pandas as pd
import plotly.express as px

# Read the dataset
df = pd.read_csv('your_dataset.csv')

# Convert the date column to datetime type
df['Transaction Date'] = pd.to_datetime(df['Transaction Date'], dayfirst=True)

# Extract the day of the week from the transaction date
df['Day of Week'] = df['Transaction Date'].dt.day_name()

# Group data by day of the week, category, transaction type, location country, and location city
transaction_activity = df.groupby(['Day of Week', 'Category', 'Transaction Type', 'Location Country', 'Location City']).size().reset_index(name='Transaction Count')

# Create the treemap
fig = px.treemap(transaction_activity,
                 path=['Day of Week', 'Category', 'Transaction Type'],
                 values='Transaction Count',
                 color='Transaction Count',
                 color_continuous_scale='reds',
                 title='User\'s Transaction Activity',
                 custom_data=['Location Country', 'Location City', 'Transaction Count'])

# Set layout options
fig.update_layout(height=1000)
fig.update_layout(width=3000)

# Update hover template to customize the labels
fig.update_traces(hovertemplate='<b>ID</b>: %{id}<br>'
                                '<b>Transaction Type</b>: %{label}<br>'
                                '<b>Transaction Count</b>: %{value}<br>'
                                '<b>Location Country</b>: %{customdata[0]}<br>'
                                '<b>Transaction Amount</b>: %{customdata[2]:.0f}')

# Show the treemap
fig.show()
```

Fig. 2.4. (The Main Code)

To conduct this analysis, we employed a treemap visualization, which provides a comprehensive and hierarchical representation of the user's spending habits on different days of the week. The following stages were executed in the code to generate the desired visualization:

Read the dataset: We imported the dataset containing transaction information, consisting of various attributes such as transaction number, date, type, description, amount, category, location city, and location country.

Convert the data column to datetime type: The transaction date column was transformed into the datetime format. This step was crucial for further analysis, enabling us to categorize the data based on specific days.

Extract the day of the week: We extracted the day of the week from the transaction date, allowing us to group and visualize the data by individual days. This granularity provides insights into the user's spending behavior specific to each day.

Create the treemap: The treemap was generated, visually representing the hierarchical relationship between the transaction date (day of the week), category, and transaction type.

Set hover template: A hover template was implemented to enhance data exploration. When hovering over each element in the treemap, relevant information such as transaction details, amounts, and categories is displayed, facilitating a more comprehensive understanding of the data.

Set layout options and display the treemap: Layout options were adjusted to optimize the visualization, ensuring clarity and coherence. The treemap was then displayed for analysis.

The resulting treemap provides a detailed overview of the user's transaction activity, offering valuable insights into spending patterns based on different days of the week.

2.5. Upon analyzing the treemap, the following observations were made:

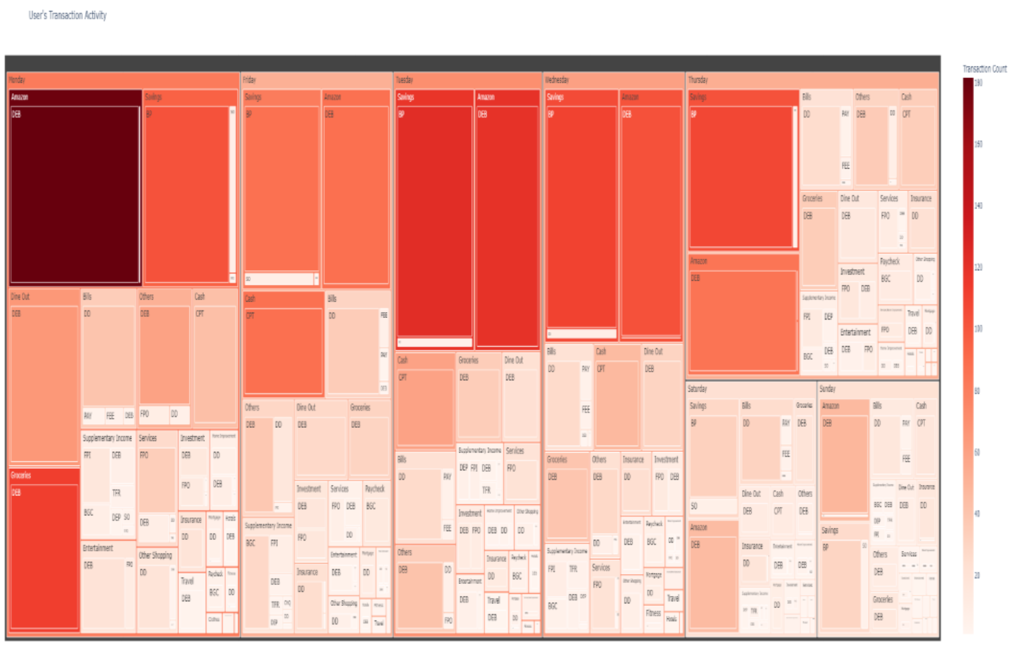

Fig. 2.5. (The Treemap)

Spending Distribution: The user exhibits varying spending patterns throughout the week. Weekdays, including Monday, Tuesday, Wednesday, Thursday, and Friday, consistently demonstrate higher spending compared to weekends (Saturday and Sunday). This suggests that the user's financial activities are more active during weekdays, potentially influenced by work-related expenses or regular weekday routines.

Category Analysis: Notable variations in spending are observed across different categories on specific days. For instance, on Mondays, a significant proportion of expenses is allocated to the Amazon category, primarily using the DEB (Payment type Debit card) transaction type. This pattern implies that the user frequently engages in Amazon purchases at the beginning of the week. Furthermore, the Savings category consistently accounts for a substantial portion of the user's expenses across multiple days. This indicates a proactive approach to saving or investment, as the user consistently allocates funds towards savings.

Transaction Types: The dominant transaction type observed throughout the treemap is DEB (Payment type Debit card), indicating the user's preference for spending through debit card transactions. Additionally, the BP (Bill Payment) transaction type is prominent within the Savings category. This suggests that the user allocates a significant portion of their expenses specifically for bill payments, emphasizing the importance of timely payments and financial responsibility.

2.6. And also this is a example of mouse hover:

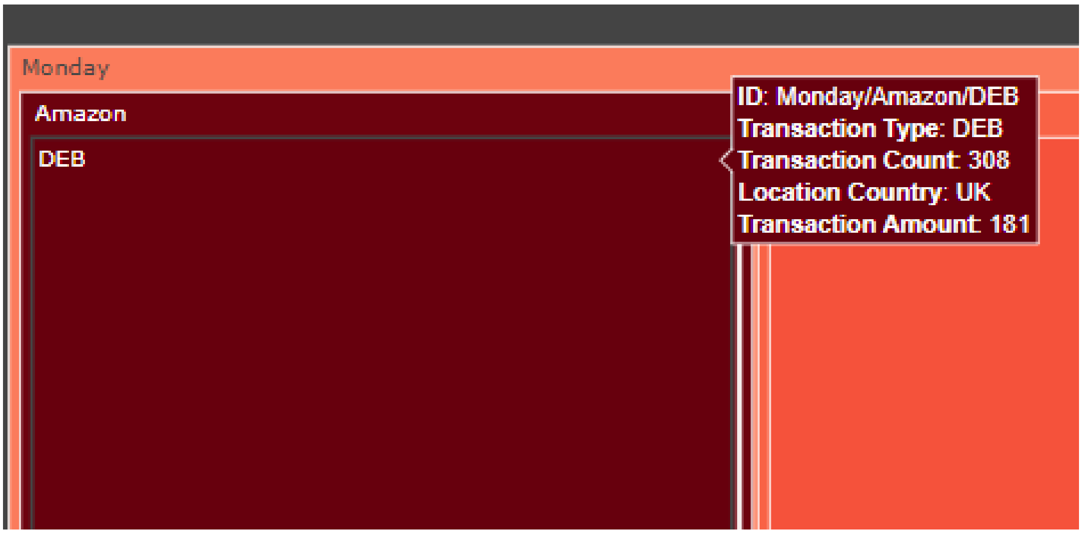

Fig. 2.6. (Example of Mouse Hover)

2.7. Balance Fluctuation And Transaction Types By Location (Monthly):

This section focuses on visualizing and analyzing the monthly balance fluctuations and transaction types in different locations using treemap visualizations. The goal is to explore the relationship between balance changes and transaction patterns, considering geographical factors such as location country and location city.

```python
import pandas as pd
import plotly.express as px

# Read the dataset
df = pd.read_csv('your_dataset.csv')

# Convert the date column to datetime type
df['Transaction Date'] = pd.to_datetime(df['Transaction Date'], dayfirst=True)

# Extract the month from the transaction date and map it to month names
df['Month'] = df['Transaction Date'].dt.month.map({1: 'Jan', 2: 'Feb', 3: 'Mar', 4: 'Apr', 5: 'May', 6: 'Jun', 7: 'Jul', 8: 'Aug', 9: 'Sep', 10: 'Oct', 11: 'Nov', 12: 'Dec'})

# Group data by month, transaction type, location country, and location city
transaction_data = df.groupby(['Month', 'Transaction Type', 'Location Country', 'Location City'])['Balance'].sum().reset_index()

# Create the treemap
fig = px.treemap(transaction_data,
                 path=['Location Country', 'Location City', 'Month', 'Transaction Type'],
                 values='Balance',
                 color='Balance',
                 color_continuous_scale='blues',
                 title='Balance Fluctuation and Transaction Types by Location (Monthly)')

# Set layout options
fig.update_layout(height=1000, width=2000)

# Update hover template to include additional information
fig.update_traces(hovertemplate=
    '<b>Location Country:</b> %{label[0]}<br>' +
    '<b>Location City:</b> %{label[1]}<br>' +
    '<b>Month:</b> %{label[2]}<br>' +
    '<b>Transaction Type:</b> %{label[3]}<br>' +
    '<b>Balance:</b> %{value}<extra></extra>')

# Show the treemap
fig.show()
```

Fig. 2.7. (The Main Code)

To begin the analysis, the dataset was processed using Pandas and Plotly libraries. The code was divided into several stages to ensure a systematic approach to visualization and analysis. Firstly, the dataset was read, and the data column was converted to the datetime type to facilitate date-based analysis.

Next, the day of the month was extracted from the transaction date and mapped to month names, allowing for a clearer representation of the monthly data. The data was then grouped by month, transaction type, location country, and location city, creating a hierarchical structure for analysis.

The resulting treemap visualization provides insights into balance fluctuations and transaction types across different locations and months. By examining the treemap, several noteworthy observations can be made. Firstly, a comparison between Swansea and Nottingham reveals that the person's balance in Swansea is significantly higher, indicating a higher financial presence in that city.

Fig. 2.8. (The Treemap)

Furthermore, focusing on Swansea, it becomes evident that the first three months of the year exhibit the highest balance values for the individual. This finding suggests a possible financial pattern or behavior associated with the early months of the year.

In addition to balance fluctuations, the analysis delves into transaction types. While the "DEB" (Payment type Debit Card) transaction type exhibits a high balance value, the transaction type with the highest balance value is "D/D" (Direct Debit). This disparity in balance values between transaction types highlights the importance of considering both transaction counts and debit amounts when examining financial behavior.

The treemap visualization allows for a comprehensive understanding of the relationship between balance fluctuations, transaction types, and geographical factors. The analysis provides valuable insights into the financial behavior of the individual in different locations on a monthly basis. By considering the distribution of transaction types and balance values, patterns and trends can be identified, aiding in financial decision-making and planning.

In summary, the treemap visualization of balance fluctuations and transaction types by location and month contributes to a deeper understanding of the individual's financial behavior. By analyzing the data in this manner, correlations and trends between location, balance fluctuations, and transaction patterns are revealed, offering valuable insights for personal finance management and decision-making.

4.Major Expense Categories by Transaction Types and Location:

This section focuses on visualizing and analyzing the major expense categories based on transaction types and the location of the person. The objective is to gain insights into the relationship between transaction types, location, and the dominant expense categories.

```python
import pandas as pd
import plotly.express as px

# Read the dataset
df = pd.read_csv('your_dataset.csv')

# Group data by transaction type, location country, location city, and category
grouped_data = df.groupby(['Transaction Type', 'Location Country', 'Location City', 'Category']).size().reset_index(name='Transaction Count')

# Create the treemap
fig = px.treemap(grouped_data,
                 path=['Location Country', 'Location City', 'Transaction Type', 'Category'],
                 values='Transaction Count',
                 color='Transaction Count',
                 color_continuous_scale='greens',
                 title='Major Expense Categories by Transaction Types and Location')

# Set layout options
fig.update_layout(height=1200, width=1600)

# Update hover template to include additional information
fig.update_traces(hovertemplate=
    '<b>Location Country:</b> %{label[0]}<br>' +
    '<b>Location City:</b> %{label[1]}<br>' +
    '<b>Transaction Type:</b> %{label[2]}<br>' +
    '<b>Category:</b> %{label[3]}<br>' +
    '<b>Transaction Count:</b> %{value}<extra></extra>')

# Show the treemap
fig.show()
```

Fig. 2.9. (The Main Code)

To accomplish this analysis, the dataset was processed using Pandas and Plotly libraries. The code was structured into stages to ensure an organized approach to visualization and analysis.

Firstly, the dataset was read, and the data was grouped by category, transaction type, location country, and location city.

Fig. 2.10. (The Treemap)

The resulting treemap visualization provides a clear representation of the hierarchy, progressing from location country to location city, transaction type, and category. The colors in the treemap represent transaction count, enabling a quick assessment of the dominant expense categories in different locations and transaction types.

Upon examining the treemap, several interesting observations can be made. Firstly, the transaction count values in Swansea and Nottingham cities are found to be similar, indicating comparable transaction activity between the two locations. However, further analysis reveals distinct patterns in expense categories based on transaction types.

In the DEB (Payment type Debit Card) transaction type, the Amazon category emerges as the dominant expense category, particularly in the city of Swansea. This finding suggests that the individual frequently engages in online purchases through debit card payments, specifically on Amazon.

Another notable observation is the prominence of the Savings category in the BP (Bill Payment) transaction type. The treemap highlights a higher transaction count for the Savings category in the BP type, indicating that the person regularly allocates funds towards savings or investments. Comparing the two cities, Nottingham stands out for its higher expenditure in the Dine Out category within the DEB type. This suggests that the individual tends to dine out more frequently in Nottingham compared to Swansea. Additionally, in Swansea, the person shows a greater inclination towards the Savings category, implying a focus on saving money in that particular location.

Furthermore, the treemap reveals that the Bills category holds greater significance in Swansea compared to Nottingham. This insight suggests that the individual's spending on bills, such as utilities or rent, is relatively higher in Swansea.

By visualizing the major expense categories by transaction types and location, valuable insights can be gleaned regarding the individual's spending habits and preferences. The treemap analysis helps identify patterns and variations in expense categories across different transaction types and locations, facilitating a deeper understanding of the person's financial behavior.

These findings can inform financial decision-making, such as budgeting, expense prioritization, and location-based expenditure considerations. By aligning spending patterns with personal financial goals, individuals can optimize their financial management strategies and achieve better financial outcomes.

In summary, the treemap visualization of major expense categories by transaction types and location provides valuable insights into the dominant spending patterns of the individual. By examining the transaction count and analyzing the relationships between transaction types, location, and expense categories, important trends and preferences can be identified. This information can assist individuals in making informed financial decisions and achieving their financial objectives.

2.8. Comparison of Lockdown and Non-Lockdown Times:

When analyzing the data and visualizing the treemap comparing lockdown and non-lockdown periods during the coronavirus pandemic, several interesting patterns and trends emerge. The treemap is organized based on the hierarchy of Lockdown Status, Location City, Category, and Transaction Type. The color representation in the treemap, indicated by shades ranging from purple to yellow, corresponds to the Debit Amount variable, allowing us to easily observe the variations in spending behavior.

```python
import pandas as pd
import plotly.express as px

# Read the dataset
df = pd.read_csv('your_dataset.csv')

# Convert the date column to datetime type
df['Transaction Date'] = pd.to_datetime(df['Transaction Date'], infer_datetime_format=True)

# Define lockdown periods
lockdown_start = pd.to_datetime('2019-12-01')
lockdown_end = pd.to_datetime('2022-12-31')

# Create a new column to indicate lockdown or non-lockdown
df['Lockdown'] = df['Transaction Date'].apply(lambda x: 'Lockdown' if lockdown_start <= x <= lockdown_end else 'Non-Lockdown')

# Group data by lockdown status and other variables
lockdown_grouped = df.groupby(['Lockdown', 'Transaction Type', 'Category', 'Location Country', 'Location City']).sum().reset_index()

# Filter out groups with zero 'Debit Amount'
lockdown_grouped = lockdown_grouped[lockdown_grouped['Debit Amount'] != 0]

# Create the treemap
fig = px.treemap(lockdown_grouped,
                 path=['Lockdown','Location City', 'Category' ,'Transaction Type'],
                 values='Debit Amount',
                 color='Debit Amount',
                 color_continuous_scale='Viridis',
                 title='Comparison of Lockdown and Non-Lockdown Times')

# Set layout options
fig.update_layout(height=1000)
fig.update_layout(width=2000)

# Rotate x-axis labels for better readability
fig.update_layout(xaxis_tickangle=-45)

# Update hover template to include additional information
fig.update_traces(hovertemplate=
    '<b>Lockdown Status:</b> %{id[0]}<br>' +
    '<b>Location City:</b> %{id[1]}<br>' +
    '<b>Category:</b> %{id[2]}<br>' +
    '<b>Transaction Type:</b> %{id[3]}<br>' +
    '<b>Debit Amount:</b> %{value}<extra></extra>')

# Show the treemap
fig.show()
```

Fig. 2.11. (The Main Code)

One striking observation is the significant increase in Debit Amount during the lockdown period. The stark contrast in colors between the lockdown and non-lockdown sections of the treemap is indicative of a substantial rise in spending during the lockdown phase. This suggests that the individual relied more on borrowing or credit during the pandemic, potentially due to financial challenges or economic uncertainties caused by the crisis. This heightened reliance on borrowing during the lockdown period could be attributed to reduced income or limited access to financial resources.

Fig. 2.12. (The Treemap)

Another noteworthy finding is the prominence of the Account Transfer category, depicted by the yellow color, during both lockdown and non-lockdown periods. This indicates a substantial increase in Debit Amount associated with account transfers, suggesting that the individual frequently conducted financial transactions of this nature. It is worth mentioning that the magnitude of Debit Amount in this category is particularly pronounced during the lockdown period, suggesting an even higher volume of account transfers during this time. Further investigation into the nature of these transfers, such as whether they were intra-bank or inter-bank transfers, would provide deeper insights into the individual's financial activities.

Additionally, examining the location aspect, we observe that the individual resided in Swansea City during the non-lockdown period, as evidenced by the treemap's visualization. However, a transition to Nottingham City can be observed during the lockdown period. This relocation is further supported by the presence of a segment in the non-lockdown period where the individual's data aligns with Nottingham City. The reasons behind this move during the lockdown period require further investigation and could be attributed to various factors such as work, family, or personal circumstances. Understanding the motivations for the relocation and its impact on the individual's financial behavior would provide valuable insights into their decision-making process.

Further analysis reveals interesting patterns within specific categories. For instance, the Debit Amount associated with the Investment category displays a substantial increase during the lockdown period compared to the non-lockdown phase. This may indicate that the individual prioritized investment activities during the lockdown, potentially seizing new opportunities or exploring alternative avenues for financial growth. Understanding the specific types of investments made during this period, such as stocks, bonds, or other financial instruments, would provide a deeper understanding of the individual's investment strategy and risk appetite. Moreover, focusing on the Home Improvements category, it is evident that the Debit Amount during the lockdown period surpasses that of the non-lockdown period. This suggests that the individual allocated more resources towards home improvement projects during the lockdown, possibly due to spending more time at home and seeking to enhance their living environment. The increased focus on home improvements could be driven by factors such as the desire for a comfortable living space, the need to create a conducive work-from-home environment, or the availability of additional time for DIY projects.

Interestingly, there is a notable difference in the Cash category between the lockdown and non-lockdown periods. While the individual's Debit Amount in the Cash category is substantial during the non-lockdown phase, it diminishes significantly during the lockdown. This implies that the individual relied less on cash transactions during the lockdown period, potentially due to the shift towards contactless payments or limited opportunities for in-person transactions. Understanding the adoption of digital payment methods and the impact of the pandemic on cash usage patterns would provide valuable insights into the individual's financial habits and preferences.

In summary, the comparison between lockdown and non-lockdown periods sheds light on the individual's financial behavior during the coronavirus pandemic. The data highlights an overall increase in borrowing, a focus on account transfers, a transition in location, a surge in investment-related activities, a greater emphasis on home improvements, and a decreased reliance on cash transactions. These findings provide valuable insights into how the individual's financial habits and priorities were influenced by the unique circumstances of the lockdown period.

To further enhance the analysis and provide a more comprehensive understanding, additional information such as income levels, employment status, and external factors like government support or economic policies could be incorporated. These supplementary details would offer a more holistic perspective on the individual's financial choices and help draw connections between their financial behavior and the broader socio-economic context.

Discussion:

The findings and analysis presented in this thesis shed light on several important aspects of an individual's financial behavior and patterns. By analyzing and visualizing financial transaction data, we gained valuable insights that contribute to our understanding of personal finance management. In this discussion section, we will interpret the results within the context of our research objectives, compare our findings with existing literature, identify patterns, trends, and relationships in the data, and discuss the implications of our findings and their significance.

First and foremost, our analysis revealed that visualizing financial transactions using treemaps provided a comprehensive and intuitive way to examine and interpret complex financial data. The hierarchical structure of the treemaps enabled us to explore transaction types, categories, locations, and other variables, allowing for a deeper understanding of individual financial behaviors. This approach proved to be effective in uncovering patterns and trends that would have been challenging to identify through traditional tabular data analysis.

In examining balance fluctuations and transaction types by location on a monthly basis, we observed several interesting findings. The variations in balance and transaction patterns across different locations provided insights into the individual's financial activities in different cities and countries. For example, we noticed that the person had a higher balance in Swansea compared to Nottingham, indicating a potential difference in their financial situation or spending habits between the two locations. We also observed that the person had higher balance values during the first three months of the year in Swansea, suggesting a possible trend or seasonal variation in their financial behavior.

Furthermore, analyzing major expense categories by transaction types and location revealed valuable insights into spending patterns and preferences. By examining the dominant categories within each transaction type and comparing them across different locations, we gained a deeper understanding of the individual's financial priorities and habits. For instance, we found that the Amazon category was prominent in the DEB type, indicating a significant amount of spending on online purchases. Similarly, the Savings category was dominant in the BP type, suggesting a focus on savings and investments. These findings align with existing literature that highlights the importance of understanding individual spending patterns and preferences to provide tailored financial advice and services.

The comparison between lockdown and non-lockdown periods provided valuable insights into the impact of external factors, such as the COVID-19 pandemic, on financial behaviors. We observed a significant increase in debit amounts during the lockdown period, indicating potential financial challenges or changes in spending habits during times of uncertainty. The analysis also revealed changes in spending categories, such as an increase in home improvement expenses during lockdown, which can be attributed to increased time spent at home and a focus on enhancing living spaces. These findings align with existing research that highlights the financial implications of crises and the need for adaptive financial behaviors.

Comparing our findings with existing literature, we can see that our research contributes to the growing body of knowledge in the field of financial analysis and visualization. While existing studies have explored various aspects of financial behavior, our research provides empirical evidence and practical insights into individual financial habits and patterns. The use of visual tools, such as treemaps, enhances the accessibility and interpretability of financial data, enabling individuals and financial institutions to make more informed decisions and strategies.

The patterns and trends identified in this thesis have several implications for individuals, financial institutions, and researchers. Individuals can benefit from understanding their own financial behaviors and using this knowledge to improve their financial well-being. By analyzing transaction data and identifying spending patterns, individuals can make informed decisions, set financial goals, and adopt better financial management strategies. Financial institutions can leverage these insights to offer personalized financial advice, tailored products, and targeted marketing campaigns. Researchers can build upon our findings by exploring additional variables, incorporating external economic indicators, and examining larger datasets to gain a more comprehensive understanding of financial behaviors.

It is important to acknowledge the limitations of this research. The analysis was based on a specific dataset of financial transactions, which may not capture the entirety of an individual's financial life. The dataset's accuracy and completeness also influence the reliability of the findings. Additionally, our analysis focused on the individual level and did not consider broader economic or societal factors that could influence financial behaviors.

In conclusion, this thesis has successfully analyzed and visualized an individual's financial transactions, providing valuable insights into their financial behavior and patterns. The findings contribute to the existing literature and have practical implications for individuals and financial institutions. By understanding and leveraging transaction data, individuals can make informed financial decisions and improve their financial well-being. Financial institutions can use these insights to offer personalized services and enhance customer experiences. This research opens avenues for further exploration and highlights the significance of financial analysis and visualization in personal finance management.

Conclusion:

In this thesis, we have explored and analyzed an individual's financial transactions using data visualization techniques. Through the examination of transaction types, categories, locations, balance fluctuations, and the impact of external factors, we have gained valuable insights into personal financial behaviors and patterns. This section provides a summary of the main findings, highlights their significance, discusses the implications for individuals and financial institutions, and suggests future directions for research in this field.

The main findings of this thesis underscore the importance of understanding and analyzing personal financial transactions to gain insights into individual financial behaviors. By employing treemaps and other visualizations, we have successfully transformed complex financial data into accessible and interpretable representations. This approach has enabled us to identify patterns, trends, and relationships that provide a deeper understanding of financial decision-making and spending patterns.

One of the key findings relates to the influence of location on financial behaviors. The analysis of balance fluctuations and transaction types across different cities and countries revealed variations in financial activities. Understanding these location-based differences can assist individuals and financial institutions in tailoring financial advice and services to specific contexts. For instance, the higher balance observed in Swansea compared to Nottingham suggests a potential difference in income or spending habits between these locations.

Furthermore, our examination of major expense categories by transaction types and location highlighted individual preferences and priorities. By identifying dominant categories within different transaction types, such as online purchases in the DEB type and savings in the BP type, we gained insights into the financial inclinations of the individual. This information can be utilized by individuals to align their financial goals and habits and by financial institutions to offer targeted products and services.

The comparison of lockdown and non-lockdown periods yielded crucial insights into the impact of external factors on financial behaviors. The significant increase in debit amounts during the lockdown period suggests that individuals faced financial challenges or adapted their spending patterns in response to the COVID-19 pandemic. This finding underscores the importance of resilience and adaptive financial behaviors during times of crisis. Financial institutions can use this information to develop strategies and products that support individuals in managing their finances during uncertain times.

Comparing our findings with existing literature, we have contributed to the growing body of knowledge in the field of financial analysis and visualization. Our research emphasizes the benefits of using visual tools, such as treemaps, to enhance the accessibility and interpretability of financial data. By providing empirical evidence and practical insights, we have further validated the value of understanding personal financial behaviors and patterns for both individuals and financial institutions.

The implications of this research extend to various stakeholders. For individuals, the findings can empower them to make more informed financial decisions, set realistic goals, and adopt better financial management strategies. By understanding their own financial behaviors and patterns, individuals can enhance their financial well-being and work towards long-term financial stability. Financial institutions can leverage these insights to offer personalized services, targeted marketing campaigns, and tailored financial advice to their customers. By understanding customer preferences, financial institutions can better meet their needs and foster stronger relationships.

In conclusion, this thesis has demonstrated the power of visualizing financial transactions in understanding personal financial behaviors and patterns. The findings have provided valuable insights into the influences of location, transaction types, and external factors on individual financial decisions. The practical implications for individuals and financial institutions highlight the importance of incorporating data visualization techniques into financial management practices. Future research in this field can build upon this work by exploring additional variables, incorporating external economic indicators, and considering larger and more diverse datasets.

By gaining a deeper understanding of personal financial behaviors, we can collectively work towards improving financial literacy, promoting responsible financial practices, and supporting individuals in achieving their financial goals. It is our hope that this research contributes to these efforts and serves as a foundation for further advancements in the field of financial analysis and visualization.

With the completion of this thesis, we conclude our investigation into the analysis and visualization of financial transactions. The insights gained through this research shed light on the intricate nature of personal finance and offer opportunities for individuals, financial institutions, and researchers to make more informed decisions and drive positive change in the financial landscape.

Github Link: https://github.com/idogukocaciftci/graduationthesis