Analyzing the Effectiveness of Bank Campaign Strategies in Promoting Financial Literacy

44517-04 Group 09 DATA WARRIORS

Northwest Missouri State University, Maryville MO 64468, USA
Group Members:

1. Sailaja Lakkakula
2. Varsha Kalvakuntla
3. Manoj Thella
4. Soujanya Kirni

1 Project Idea:

This research will assess the performance of a bank's current financial literacy efforts and make recommendations for future campaigns. The study aims to uncover critical factors influencing campaign efficacy and produce actionable insights to improve the impact of financial literacy initiatives through the analysis of demographic data, campaign engagement indicators, and financial literacy scores.

2 Tools and Technologies:

1. Tableau
2. Python
3. PySpark
4. Apache Spark
5. Matplotlib
6. Jupyter Notebook

3 Architecture Diagram and Summary

- **Financial Literacy Data Dataset:**This is the primary dataset containing information about campaign participants, including demographic data and financial literacy scores.
- **Data Preprocessing:** This step involves cleaning and preparing the dataset for analysis, including handling missing values, encoding categorical variables, and standardizing numerical features.
- **Demographic Data:** Demographic data includes information about participants' age, gender, income level, education level, and employment status

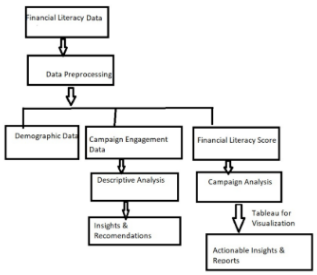

Fig. 1. Campaign Data flow

- **Campaign Engagement Data:** This data captures participants' interactions with the bank's campaign, such as attendance at workshops, usage of online resources, and engagement on social media.
- **Financial Literacy Scores:**These scores indicate participants' levels of financial literacy, which may be assessed through knowledge assessments or self-reported measures.
- **Financial Literacy Scores:**These scores indicate participants' levels of financial literacy, which may be assessed through knowledge assessments or self-reported measures.
- **Descriptive Analysis:**This phase involves summarizing and exploring the demographic characteristics of participants, as well as their engagement with campaign activities
- **Campaign Effectiveness Analysis:**Here, the effectiveness of various campaign strategies in promoting financial literacy is analyzed, including comparisons of engagement levels and financial literacy outcomes across different channels.
- **Insights and Recommendations:**Based on the analysis results, actionable insights and recommendations are generated to optimize future campaign strategies and improve financial literacy promotion efforts.
- **Actionable Insights Report:** This is the final output of the analysis, presenting key findings, recommendations, and actionable insights for stakeholders to inform decision-making and strategy development.

  By leveraging data-driven insights and targeted optimization strategies, this project aims to enhance the effectiveness of financial literacy campaigns conducted by the bank, ultimately empowering individuals to make informed financial decisions and improve their financial well-being.
- **Producing Filtered Data to Another Kafka Topic:** Finally, the processed and filtered tweet data is produced into another Kafka topic. This topic serves as an output for the refined tweet data, which can then be consumed by downstream applications or services for further analysis, visualization, or storage.

4 Project Goals

1. Analyze Monthly Campaign Effectiveness by Marital Status and Day of the Week.
   (a) Analyze monthly campaign effectiveness by marital status and day of the week.
   (b) Utilize response rates, conversion rates, and engagement levels for insights.
   (c) Identify patterns and trends across marital statuses and weekdays.
   (d) Optimize campaign scheduling and targeting strategies based on analysis.
   (e) Improve overall campaign performance by leveraging insights.

2. Analyze the effectiveness of the bank's campaign by determining factors that influence whether a customer participates in the campaign or not.
   (a) Analyze factors influencing customer participation in the campaign.
   (b) Categorize customers by participation status and job.
   (c) Calculate total participation counts for each job category.
   (d) Visualize job counts and total participation to identify trends.
   (e) Utilize insights to optimize campaign targeting and outreach strategies.

3. Build predictive models to forecast whether a customer will participate in the campaign based on their past interactions and demographic/financial attributes.
   (a) Choose suitable predictive modeling techniques such as logistic regression or decision trees.
   (b) Split data into training and testing sets for model evaluation.
   (c) Train predictive model using Tableau's integration with R or Python.
   (d) Evaluate model performance using appropriate metrics like accuracy or ROC curve.
   (e) Optimize model hyperparameters and deployment for real-time predictions

4. Optimize future marketing campaigns by identifying the most effective channels, timing and contact duration.
   (a) Import relevant dataset attributes into Tableau.
   (b) Select and analyze contact, month, day of week, and campaign duration.
   (c) Explore effectiveness of channels, timing, and contact duration.
   (d) Identify most effective marketing channels and optimal timing for campaigns.
   (e) Optimize future marketing strategies based on analysis findings.

5. Determine how economic conditions influence customer behavior and campaign effectiveness.
   (a) Analyze how economic indicators influence customer behavior and campaign success.

   (b) Explore relationship between employment variation rate and consumer price index.
   (c) Utilize Spark DataFrame for data loading and analysis.
   (d) Handle missing values and perform exploratory data analysis.
   (e) Visualize distribution of economic indicators to gain insights.

6. Assess the long-term impact of campaign participation on customer behavior, such as account activity, deposits, loans.
   (a) Assess long-term impact of campaign participation on customer behavior.
   (b) Track customer actions such as account activity and financial decisions post-campaign.
   (c) Understand influence of campaign participation on customer relationships.
   (d) Gain insights to optimize future campaign strategies for better retention and engagement.
   (e) Inform decision-making based on analysis of long-term customer behavior.

7. Measure customer retention rates and lifetime value to determine the overall success of the campaign strategy.
   (a) Measure customer retention rates and lifetime value.
   (b) Categorize customers into age groups.
   (c) Calculate total interactions and returning customers for each age group.
   (d) Determine retention rates and lifetime value by age group.
   (e) Analyze campaign success based on retention rates and lifetime value.

8. Determine the optimal number of contacts to maximize participation while minimizing customer fatigue.
   (a) Gather data on customer interactions and demographic information.
   (b) Clean and preprocess data to ensure accuracy.
   (c) Explore relationship between contact frequency and participation rates.
   (d) Segment customers based on demographic attributes.
   (e) Develop predictive models and conduct optimization analysis for optimal contact frequency.

5 Implementation of Project Goals

5.1 Goal 1:

Analyze Monthly Campaign Effectiveness by Marital Status and Day of the Week.

```python
from pyspark.sql import SparkSession

# Initialize SparkSession
spark = SparkSession.builder \
    .appName("CampaignContactAnalysis") \
    .getOrCreate()

# Load the dataset into a DataFrame
historical_data = spark.read.csv("C:/Users/SS63147/Desktop/bank-additional-full (1).csv", header=True, inferSchema=True)

# Register the DataFrame as a temporary view
historical_data.createOrReplaceTempView("campaign_data")

# Execute the SQL query to analyze contact patterns
result = spark.sql('''
    SELECT day_of_week, contact, month, AVG(duration) AS avg_duration, COUNT(*) AS contact_count FROM campaign_data GROUP BY day_of_week, contact, month ORDER BY  day_of_week,
    month limit 15
''')

# Show the result
result.show()

# Stop SparkSession
spark.stop()
```

Fig. 2. Goal 1 Implementation

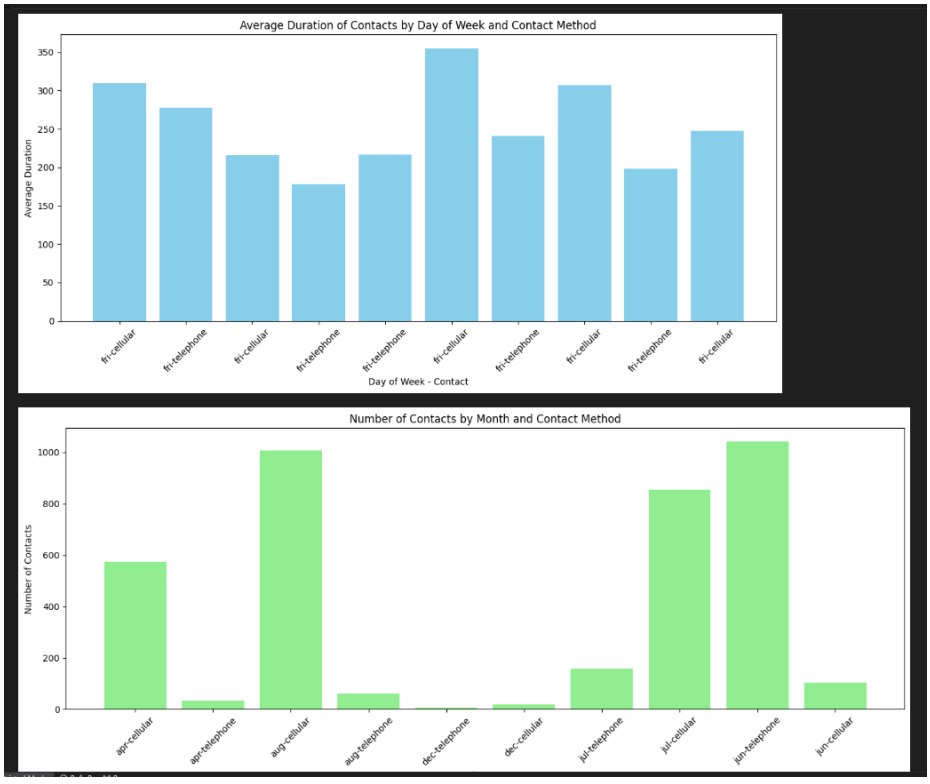

Fig. 3. Goal 1 Output

5.2 Goal 2:

Analyze the effectiveness of the bank's campaign by determining factors that influence whether a customer participates in the campaign or not.

```python
from pyspark.sql import SparkSession
from pyspark.sql.functions import when, col, sum

# Create a SparkSession
spark = SparkSession.builder \
    .appName("Campaign Participation Analysis") \
    .getOrCreate()

# Load the dataset
df = spark.read.csv("C:\\Users\\s559494\\OneDrive - nwmissouri.edu\\Documents\\NWMSU\\Subjects\\Spring 24\\Big Data\\Assignments\\bank-additional-full.csv",
                    header=True, inferSchema=True)

# Add a column for participation status
df = df.withColumn("participation_status", when(col("campaign") == 0, "Not Participated").otherwise("Participated"))

# Group by job and participation status, and count the occurrences
result = df.groupBy("job") \
    .pivot("participation_status") \
    .count() \
    .na.fill(0)  # Fill null values with 0

# Calculate total participation counts
total = result.groupBy().agg(
    sum("Not Participated").alias("Total Not Participated"),
    sum("Participated").alias("Total Participated")
)

# Print Total Section
print("Total Section:")
total.select("Total Not Participated", "Total Participated").show()

# Print Job Counts Section
print("Job Counts Section:")
result.show()

# Visualize the results
import matplotlib.pyplot as plt

# Extract data for visualization
job_counts = result.toPandas()
total_counts = total.toPandas().iloc[0]

# Plot Job Counts
plt.figure(figsize=(10, 6))
job_counts.plot(kind='bar', x='job', stacked=True, colormap='viridis')
plt.title('Job Counts by Participation Status')
plt.xlabel('Job')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Participation Status')
plt.tight_layout()
plt.show()
```

Fig. 4. Goal 2 Implementation

Total Section:

Total Not Participated	Total Participated
10795	30393

Job Counts Section:

job	Not Participated	Participated
management	795	2129
retired	435	1285
unknown	68	262
self-employed	355	1066
student	224	651
blue-collar	2437	6817
entrepreneur	427	1029
admin.	2738	7684
technician	1722	5021
services	1078	2891
housemaid	276	784
unemployed	240	774

Fig. 5. Goal 2 Output

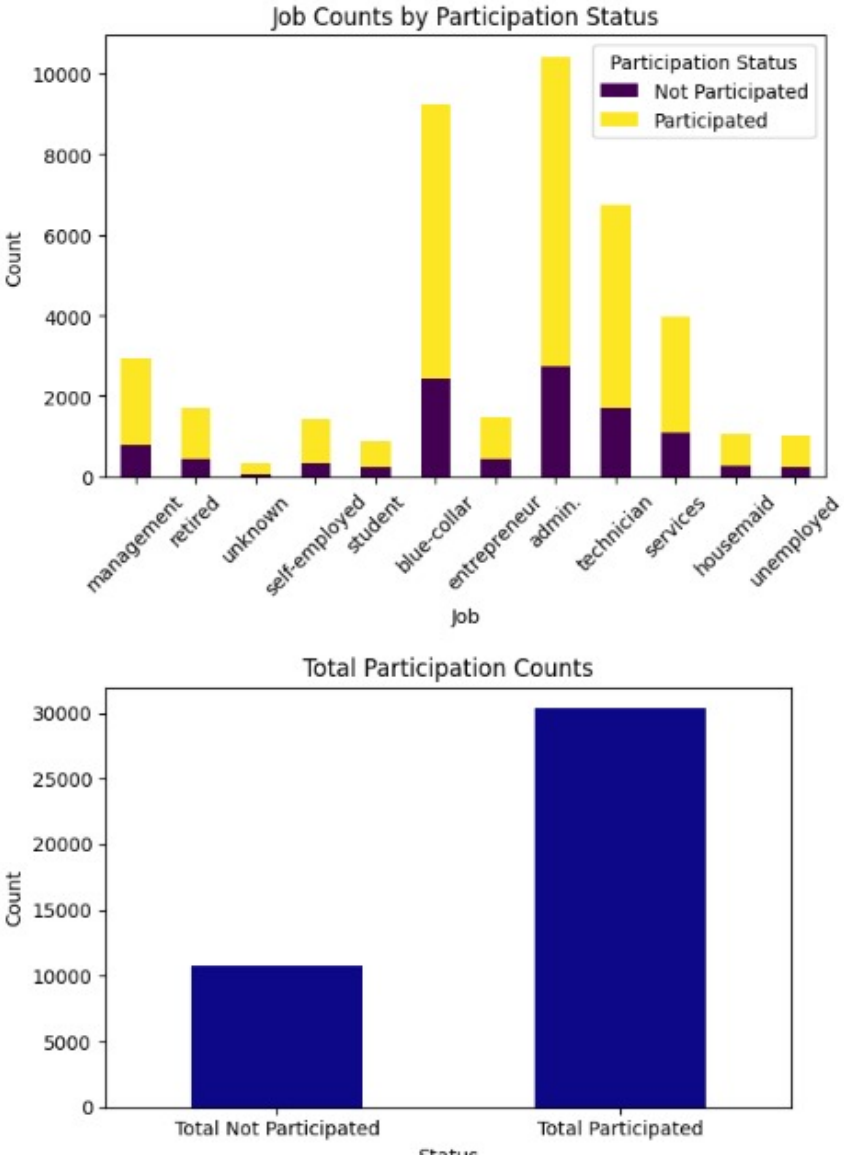

Fig. 6. Goal 2 Output

5.3 Goal 3:

Build predictive models to forecast whether a customer will participate in the campaign based on their past interactions and demographic/financial attributes.

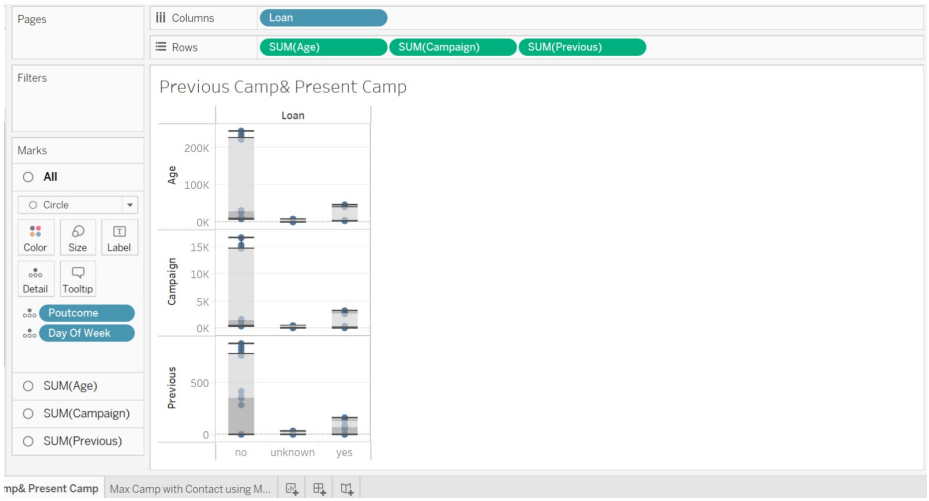

Fig. 7. Goal 3 Output

5.4 Goal 4:

Optimize future marketing campaigns by identifying the most effective channels, timing and contact duration.

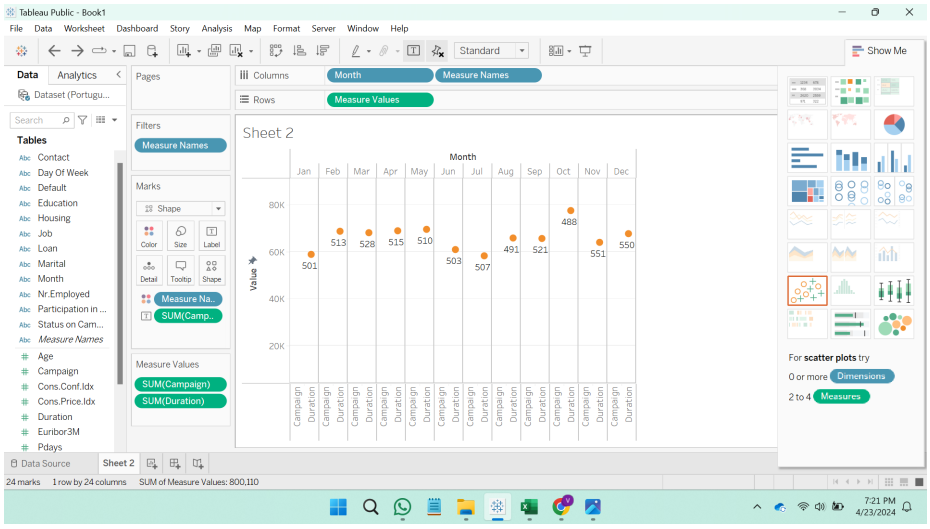

Fig. 8. Goal 4 Implementation

5.5 Goal 5:

Determine how economic conditions influence customer behavior and campaign effectiveness.

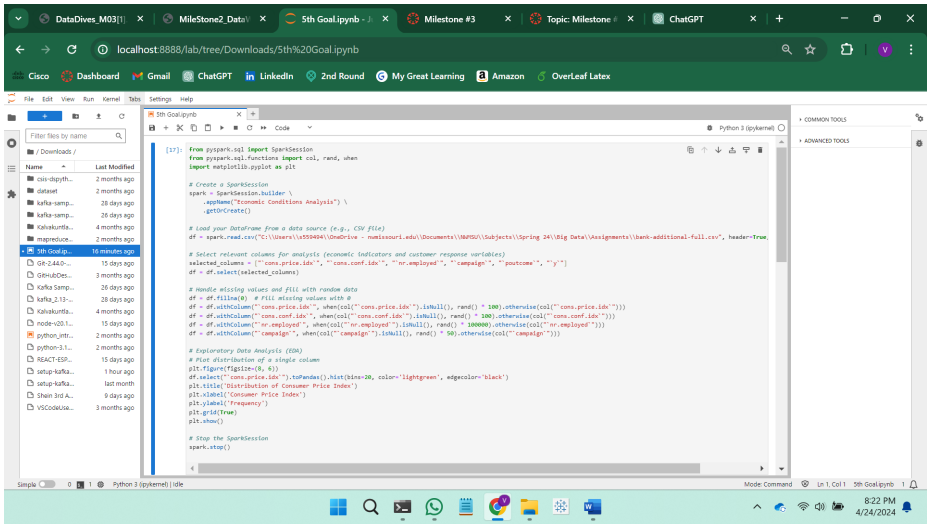

Fig. 9. Goal 5 Implementation

Fig. 10. Goal 5 output

5.6 Goal 6:

Assess the long-term impact of campaign participation on customer behavior, such as account activity, deposits, loans.

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import when, col, sum

# Create a SparkSession
spark = SparkSession.builder \
    .appName("Campaign Participation Analysis") \
    .getOrCreate()

# Load the dataset
df = spark.read.csv("C:\\Users\\s559494\\OneDrive - nwmissouri.edu\\Documents\\NWMSU\\Subjects\\Spring 24\\Big Data\\Assignments\\bank-additional-full.csv",
                    header=True, inferSchema=True)

# Add a column for participation status
df = df.withColumn("participation_status", when(col("campaign") == 0, "Not Participated").otherwise("Participated"))

# Group by job and participation status, and count the occurrences
result = df.groupBy("job") \
    .pivot("participation_status") \
    .count() \
    .na.fill(0)  # Fill null values with 0

# Calculate total participation counts
total = result.groupBy().agg(
    sum("Not Participated").alias("Total Not Participated"),
    sum("Participated").alias("Total Participated")
)

# Print Total Section
print("Total Section:")
total.select("Total Not Participated", "Total Participated").show()

# Print Job Counts Section
print("Job Counts Section:")
```

Fig. 11. Goal 6 Implemetation

campaign_participation	total_account_activity	total_deposits	total_loans
No	2403831	NULL	1584
Yes	6130351	NULL	4032

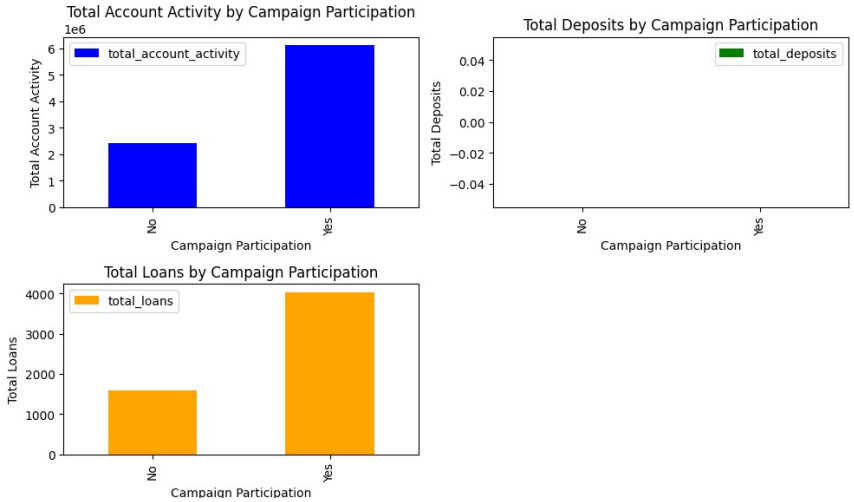

Fig. 12. Goal 6 output

5.7 Goal 7:

Measure customer retention rates and lifetime value to determine the overall success of the campaign strategy.

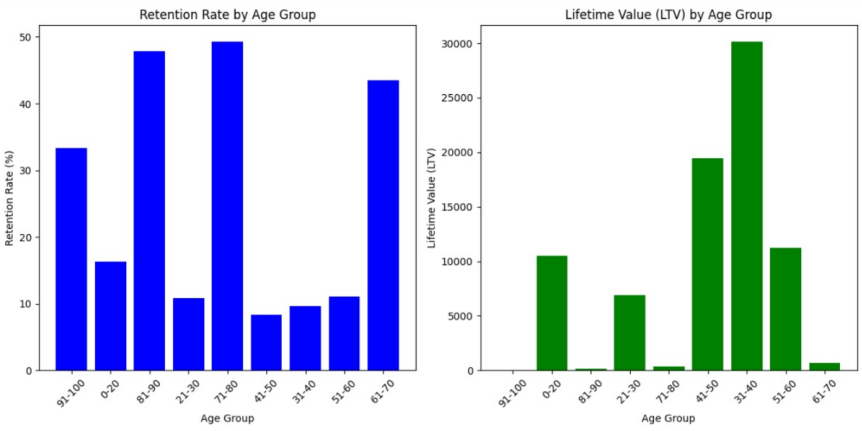

Fig. 13. Goal 7 output

5.8 Goal 8:

Determine the optimal number of contacts to maximize participation while minimizing customer fatigue.

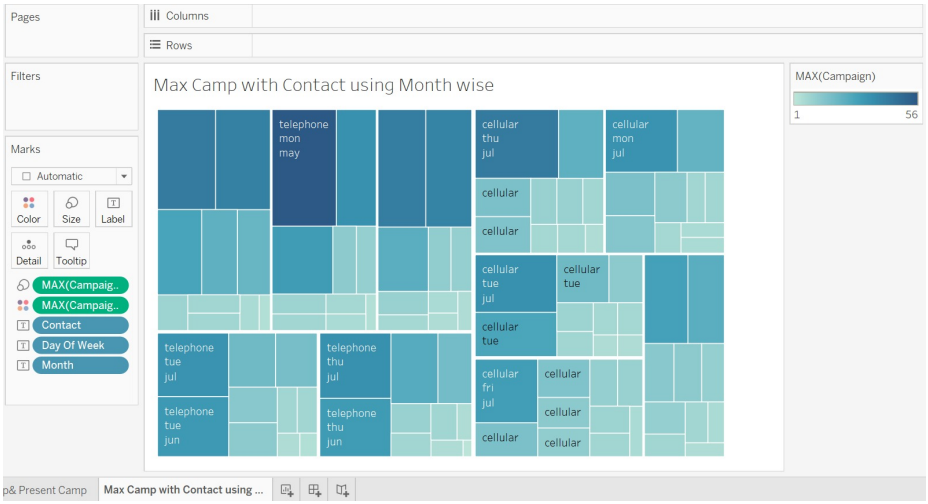

Fig. 14. Goal 8 output

6 Conclusion:

Our project proposal aims to harness the power of data analytics to drive informed decision-making in our bank's marketing campaigns. By leveraging advanced tools and technologies like Tableau, Apache Spark, Kafka, and Python, our team is poised to delve deep into the campaign dataset and extract actionable insights.Through segmentation analysis, we intend to gain a deeper understanding of our customer base, allowing us to tailor our marketing efforts more effectively. Analyzing campaign effectiveness will help us identify key factors influencing customer participation, while predictive modeling will enable us to anticipate future behaviors.Optimization of marketing campaigns, informed by data-driven insights, will be crucial in maximizing participation and minimizing customer fatigue. Additionally, by evaluating economic conditions and assessing long-term impacts, we aim to ensure the sustainability and success of our campaign strategies.Ultimately, our goal is to enhance customer retention rates, measure lifetime value, and determine the optimal contact strategies to achieve our campaign objectives. With a comprehensive approach to data analysis and a focus on strategic decision-making, we are confident in the success of our project and its potential to drive positive outcomes for our bank.

7 Citations

7.1 Github url :

https://github.com/ManojThella/BigDataProject