Tableau Insight: : Illuminating Insights From Uber Expeditionary Analytics

Provide background information on Uber and its data analytics division, highlighting the importance of data-driven decision-making in the company's operations. Outline the specific goals of the internship project, such as analyzing customer behavior patterns, optimizing operational efficiency, and improving service quality. Review existing literature on data analytics in the transportation industry, focusing on similar case studies and best practices. Detail the sources of data utilized, such as Uber's internal databases, customer feedback surveys, and market research reports. Explain the steps taken to clean, organize, and prepare the raw data for analysis, including handling missing values and outliers. Provide actionable recommendations for Uber based on the insights derived from the data analysis, including strategies for improving customer satisfaction and operational efficiency.

Real-Time Scenarios:

Scenario 1: A Case Study of Uber Expeditionary Analytics":

In a real-time scenario Data-Driven Decision Making in the Transportation Industry"* by Smith, J. et al. (Year)

- This paper explores the significance of data-driven decision-making in the transportation sector, highlighting the role of data analytics in improving operational efficiency and customer satisfaction.. This review article provides an overview of various customer segmentation methods used in ride-sharing services, including clustering algorithms and machine learning techniques, and their applications in enhancing service quality and personalized experiences. This study investigates the use of predictive modeling techniques, such as time series analysis and machine learning algorithms, for demand forecasting in transportation services. It discusses the challenges and opportunities of applying these methods to improve service reliability and resource allocation.

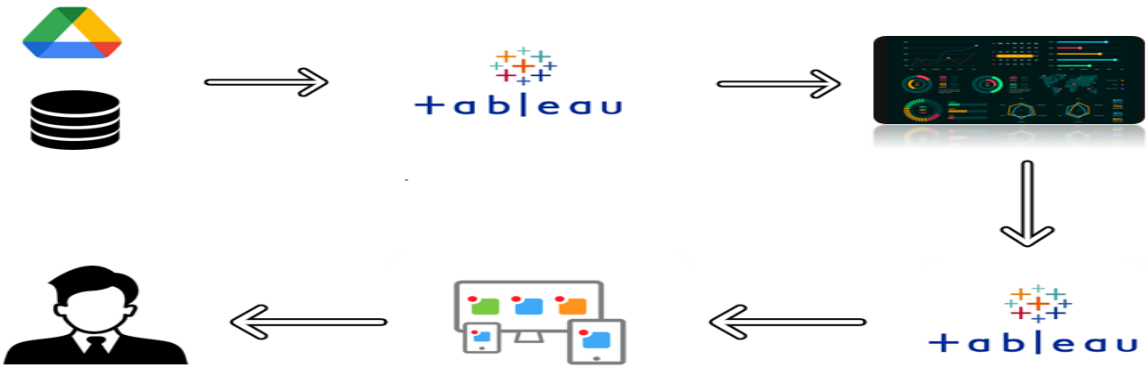
Scenario 2: *Data-Driven Decision Making:

Theoretical frameworks such as the DIKW (Data-Information-Knowledge-Wisdom) pyramid can be applied to illustrate how raw data collected by Uber can be transformed into actionable insights to enhance customer experience. This framework highlights the progression from data to wisdom, emphasizing the importance of analytics in extracting meaningful knowledge from vast datasets.

Scenario 3: Continuous Improvement:

Theoretical concepts from quality management, such as the Plan-Do-Check-Act (PDCA) cycle and continuous improvement methodologies, can be applied to ensure that data-driven initiatives at Uber are iterative and adaptive. By regularly monitoring key performance indicators and soliciting feedback from customers, Uber can identify areas for improvement and implement targeted interventions to enhance the customer experience over time.

Technical Architecture:



Project Flow

To accomplish this, we have to complete all the activities listed below,

- Data Collection & Extraction from Database
  - Collect the dataset,
  - Connect data with Tableau
- Data Preparation
  - Prepare the Data for Visualization
- Data Visualizations
  - No of Unique Visualizations
- Dashboard
  - Responsive and Design of Dashboard
- Story
  - No of Scenes of Story
- Performance Testing
  - Amount of Data Loaded
  - Utilization of Data Filters
  - No of Calculation Fields
  - No of Visualizations/ Graphs
- Web Integration
  - Dashboard and Story embed with UI With Flask
- Project Demonstration & Documentation
  - Record explanation Video for project end to end solution
  - Project Documentation-Step by step project development procedure

Milestone 1: Data Collection & Extraction from Database

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes and generate insights from the data.

Activity 1.1: Understand the data

Data contains all the meta information regarding the columns described in the CSV files.

Column Description of the Datase:

1.Data Collection: Gather data from various sources. This might include:

- Uber's own data analytics platform (if available)
- Publicly available data sources such as government transportation data, census data, etc.
- Surveys or interviews with Uber drivers or users
- Third-party data providers that specialize in transportation or market analysis

2.Data Extraction: Once you've identified your data sources, you'll need to extract the relevant data. This could involve:

- Web scraping for publicly available data
- API access if Uber provides an API for accessing certain data
- Database queries if you're accessing data from a structured database
- Manual data entry or transcription if dealing with non-digital sources

3.Data Cleaning and Preparation: Raw data often needs cleaning and preprocessing before it can be analyzed. This might involve:

- Removing duplicates or erroneous data points
- Handling missing data through imputation or deletion
- Standardizing formats and units across different datasets
- Encoding categorical variables

4.Data Analysis: With your cleaned and prepared data, you can now perform your analysis. This could involve various statistical techniques, machine learning algorithms, or data visualization methods, depending on your objectives and the nature of your data.

Interpretation and Reporting: Once you've analyzed the data, you'll need to interpret the results and draw conclusions Throughout this process, it's important to consider ethical considerations related to data privacy, consent, and responsible data usage, especially when dealing with sensitive information about users or drivers.

Milestone 2: Data Preparation

Activity 1: Prepare the Data for Visualization

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for creating visualizations to gain insights into the performance and efficiency.

Explanation video link 1: Data Loading

https://drive.google.com/file/d/12W1BDNTkyGAE55V8EWtY9ubID7Ms9m6W/view?usp=drivesdk

Milestone 3: data visualization+

Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

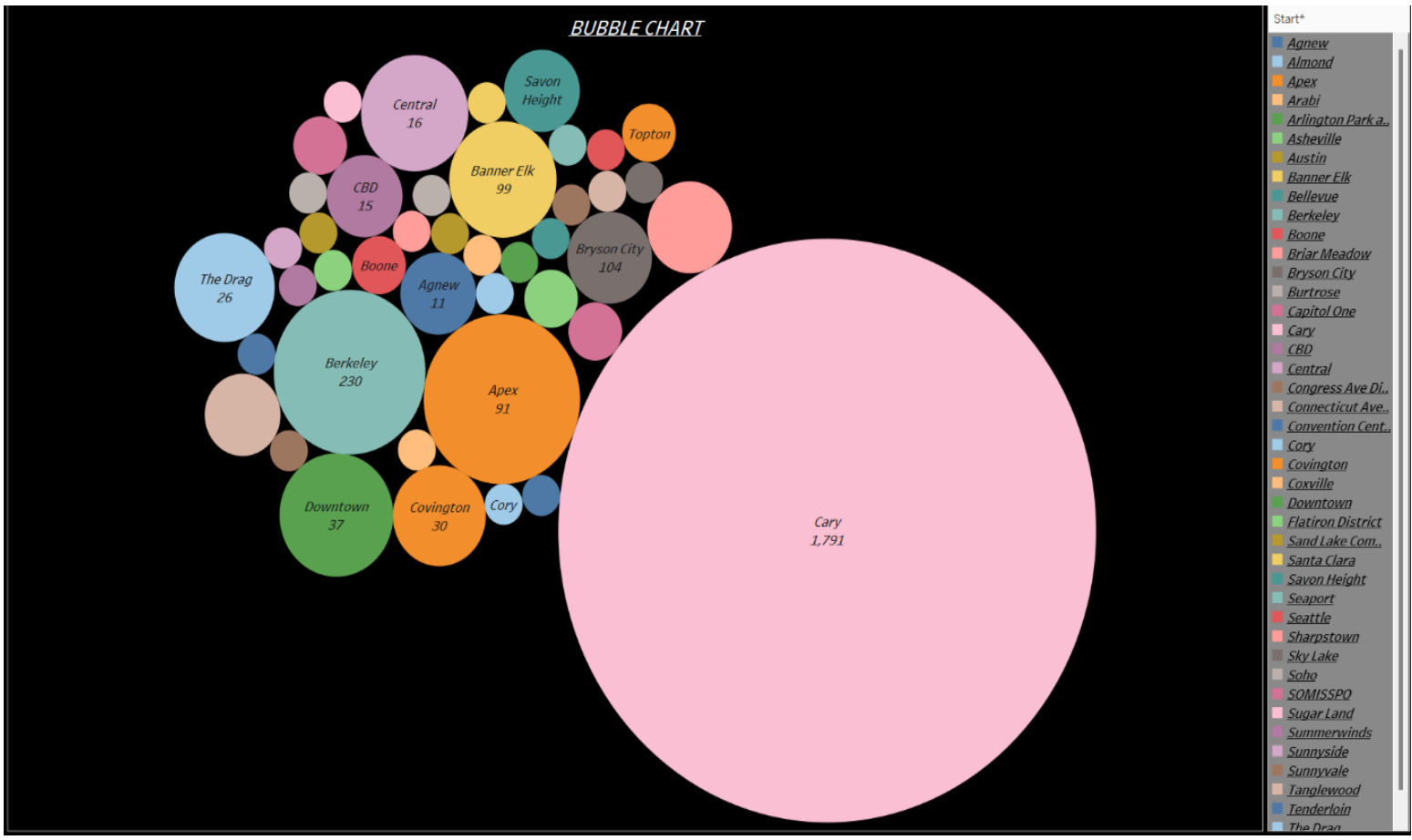
Activity 1: No of Unique Visualizations

The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyse the performance and efficiency of Radisson Hotels include bar charts, line charts, heat maps, scatter plots, pie charts, Maps etc. These visualizations can be used to compare performance, track changes over time, show distribution, and relationships between variables, breakdown of revenue and customer demographics, workload, resource allocation and location of hotels.

Activity 1.1: total number miles for the date of years

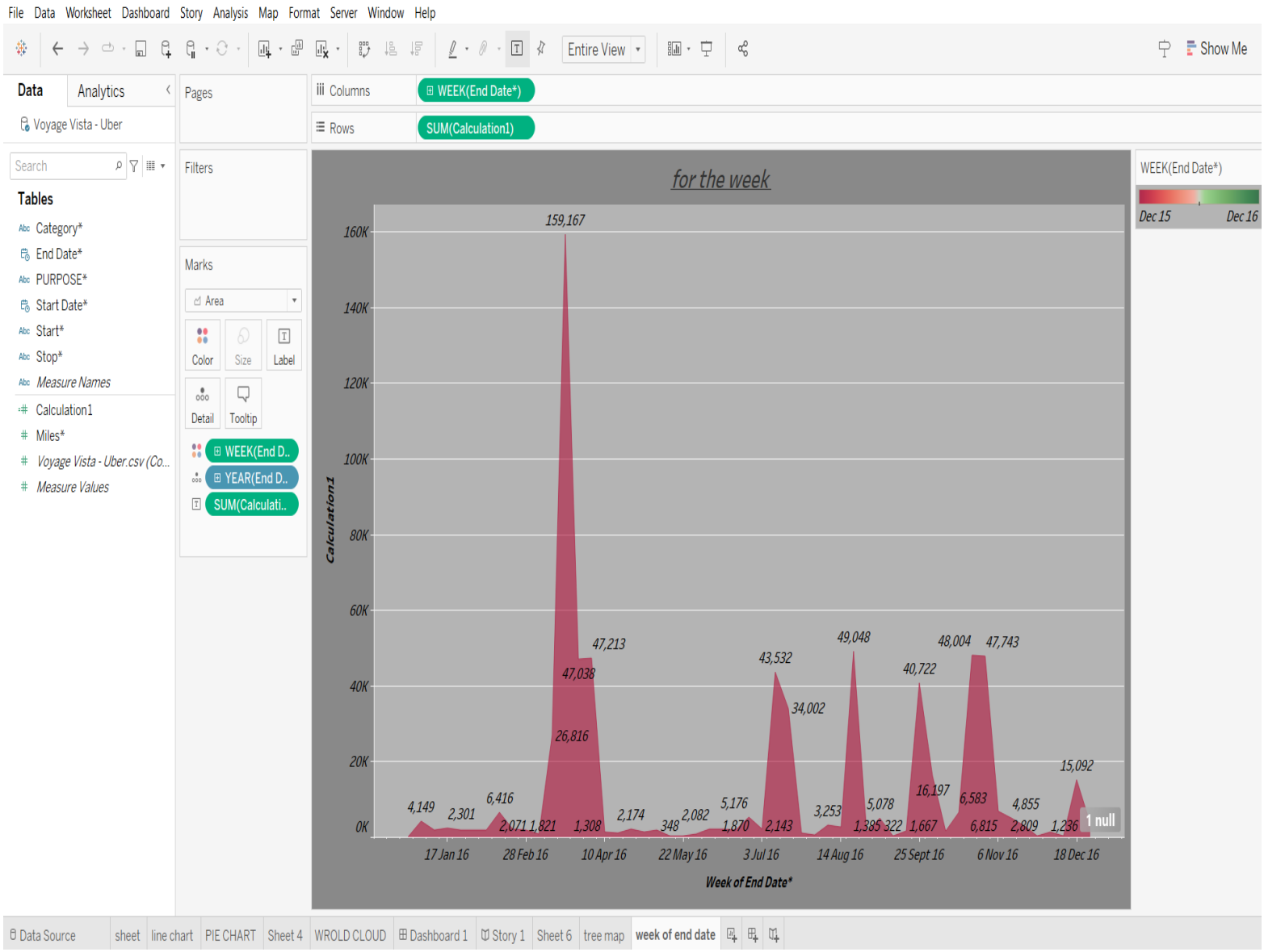
Explanation video link:

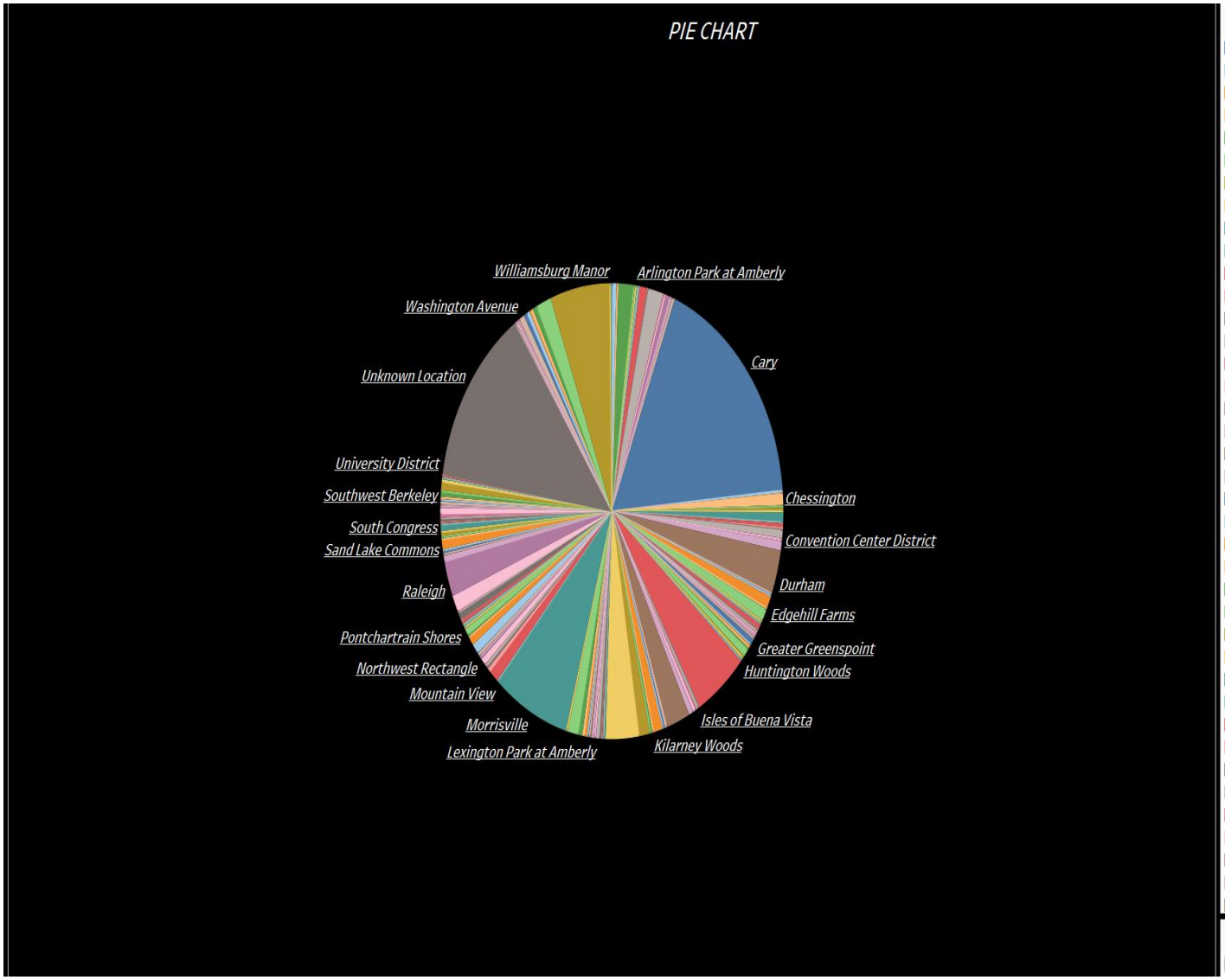
https://drive.google.com/file/d/12W1BDNTkyGAE55V8EWtY9ubID7Ms9m6W/view?usp=drivesdk



Activity 1.3: Explanation video link: total count of the vista uber by the stop uber annalyst:

https://drive.google.com/file/d/12W1BDNTkyGAE55V8EWtY9ubID7Ms9m6W/view?usp=drivesdk




PIE CHART
Williamsburg Manor
Arlington Park at Amberly
Washington Avenue
Cary
Unknown Location
University District
Southwest Berkeley
Chessington
South Congress
Convention Center District
Sand Lake Commons
Durham
Raleigh
Edgehill Farms
Pontchartrain Shores
Greater Greenspoint
Northwest Rectangle
Huntington Woods
Mountain View
Isles of Buena Vista
Morrisville
Kilarney Woods
Lexington Park at Amberly
Stop*
Null
Agnew
Alief
Almond
Apex
Arabi
Arlington
Arlington Park a..
Arts District
Asheville
Banner Elk
Bay Farm Island
Bellevue
Berkeley
Boone
Briar Meadow
Bryson City
Burtrose
Bywater
Capitol One
Cary
CBD
Cedar Hill
Central
Chalmette
Chapel Hill
Chessington
College Avenue
Colombo
Columbia Heigh..
Congress Ave Di..
Convention Cent..
Cory
Covington
Coxville
Daytona Beach
Depot Historic D..
Downtown
Durham
SUM(Miles*)
24,409

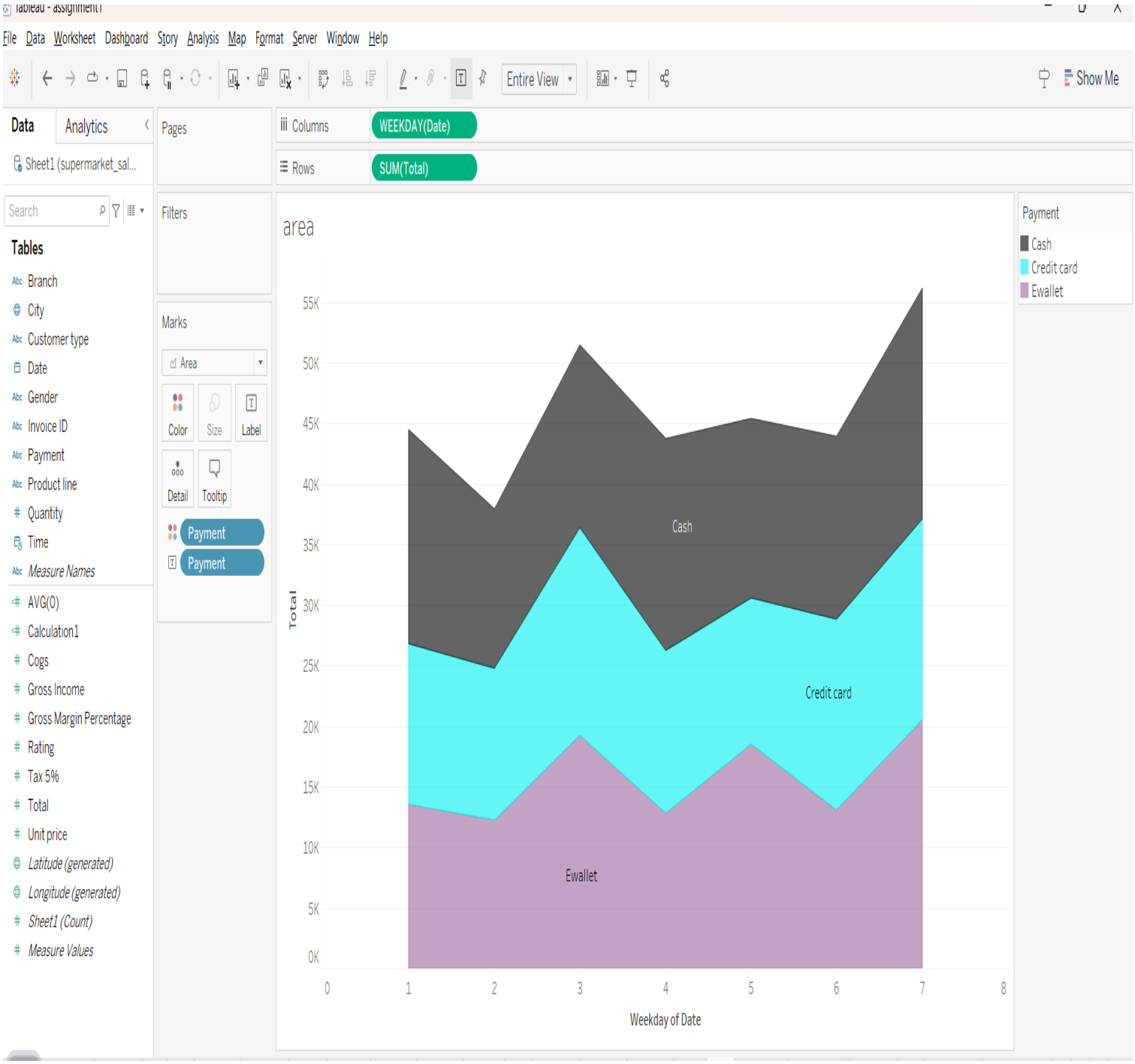
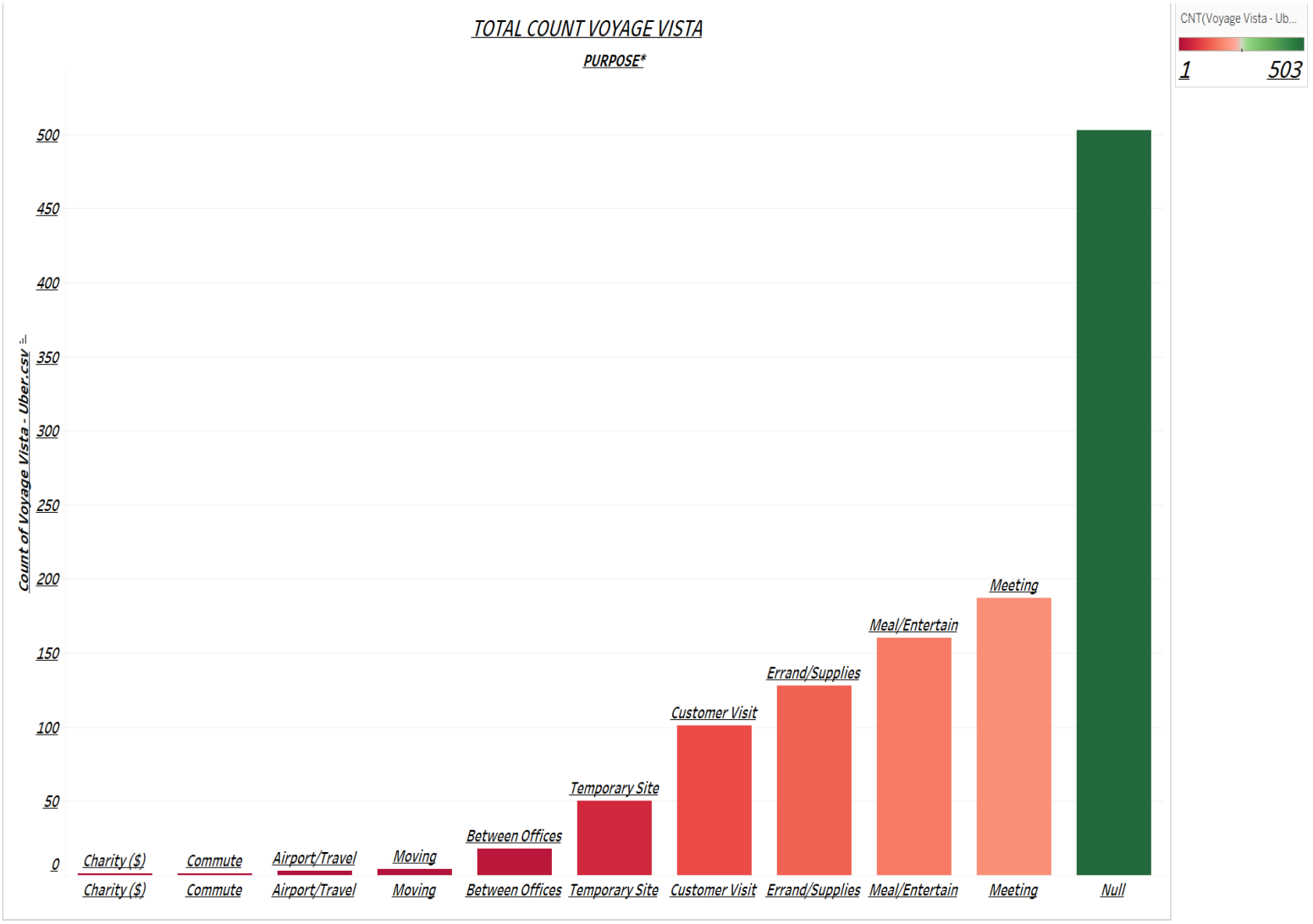


Tableau - assignment1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
Sheet1 (supermarket_sal...
Pages
Columns
WEEKDAY(Date)
Rows
SUM(Total)
Search
Filters
Tables
Branch
City
Customer type
Date
Gender
Invoice ID
Payment
Product line
Quantity
Time
Measure Names
AVG(0)
Calculation1
Cogs
Gross Income
Gross Margin Percentage
Rating
Tax 5%
Total
Unit price
Latitude (generated)
Longitude (generated)
Sheet1 (Count)
Measure Values
Marks
Area
Color
Size
Label
Detail
Tooltip
Payment
Payment
area
Payment
Cash
Credit card
Ewallet
55K
50K
45K
40K
35K
30K
25K
20K
15K
10K
5K
0K
Total
Cash
Credit card
Ewallet
0
1
2
3
4
5
6
7
8
Weekday of Date

Activity 1.4: *number of the voyage vista for the total number purpose:*

Explanation video link

https://drive.google.com/file/d/12W1BDNTkyGAE55V8EWtY9ubID7Ms9m6W/view?usp=drivesdk

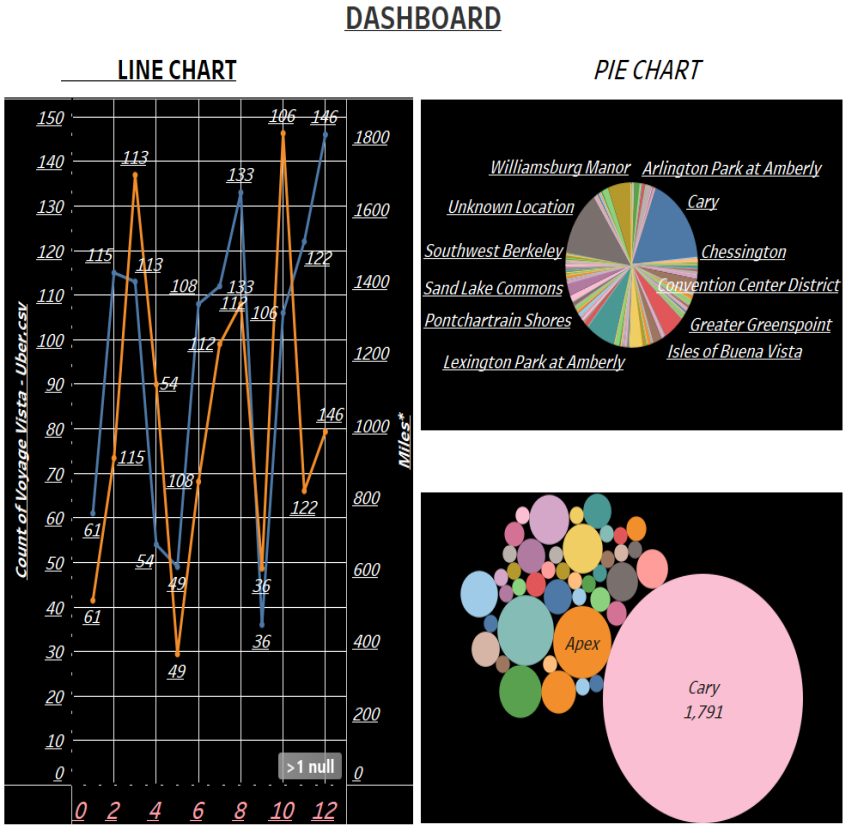


Milestone 4: Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data, and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Activity :1- Responsive and Design of Dashboard

https://drive.google.com/file/d/12W1BDNTkyGAE55V8EWtY9ubID7Ms9m6W/view?usp=drivesdk



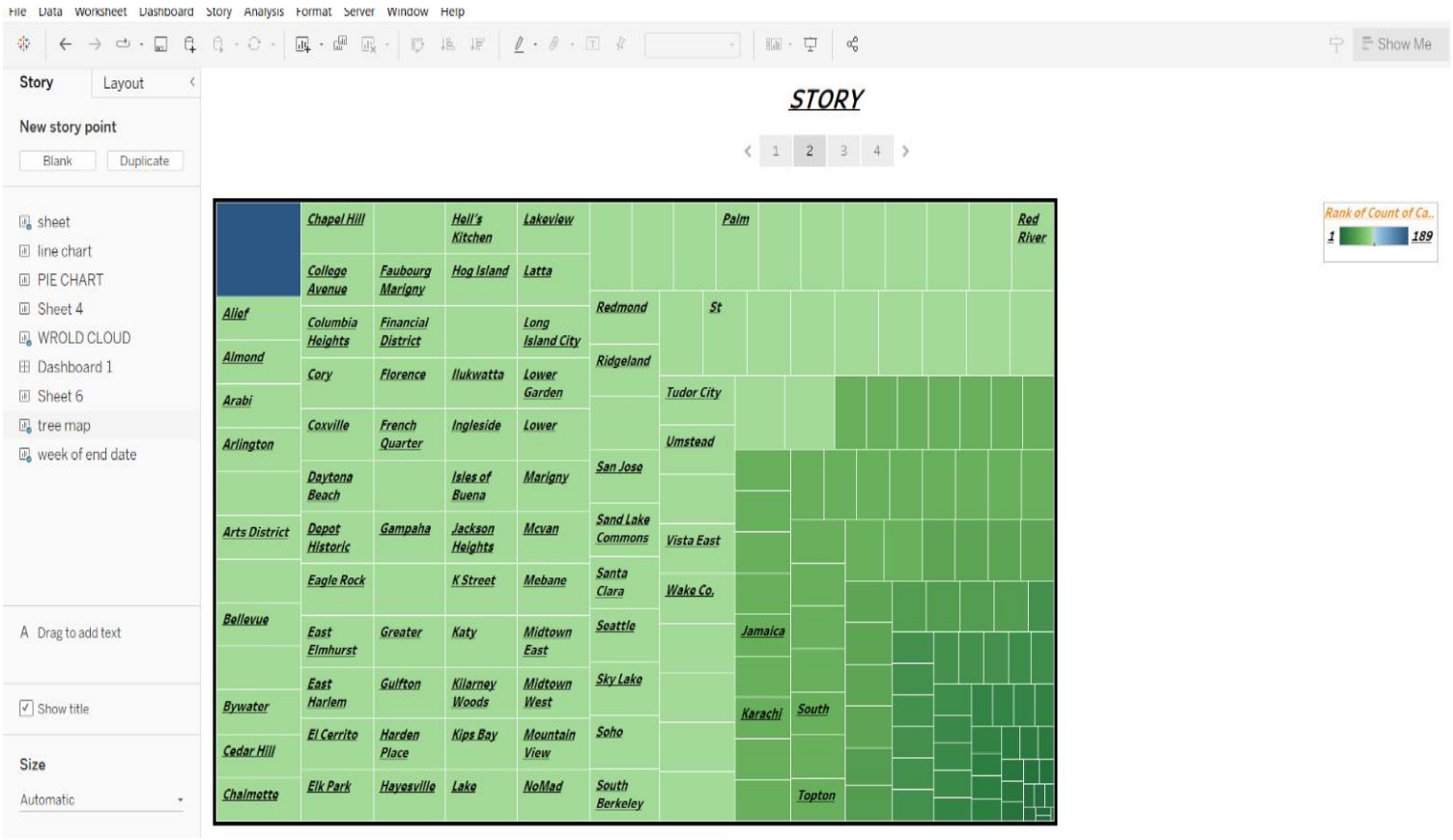
Milestone 5: story

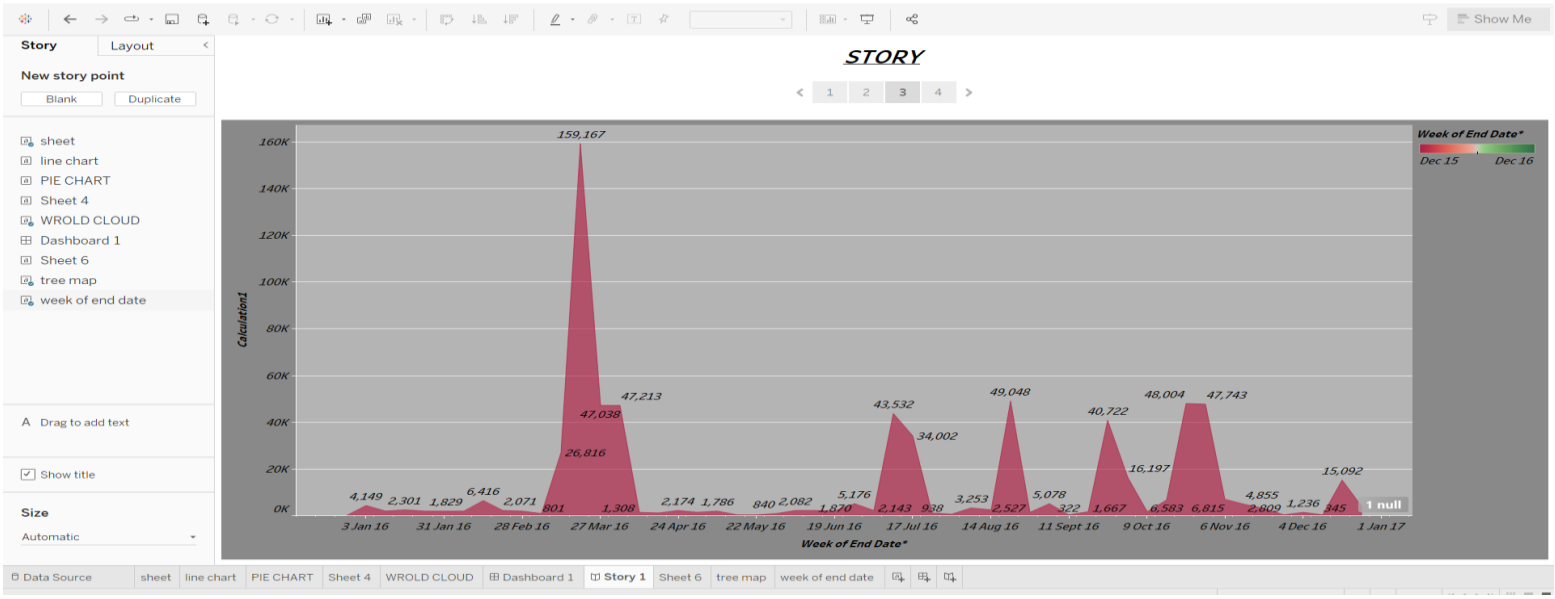
A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

Activity 1: No of Scenes of Story

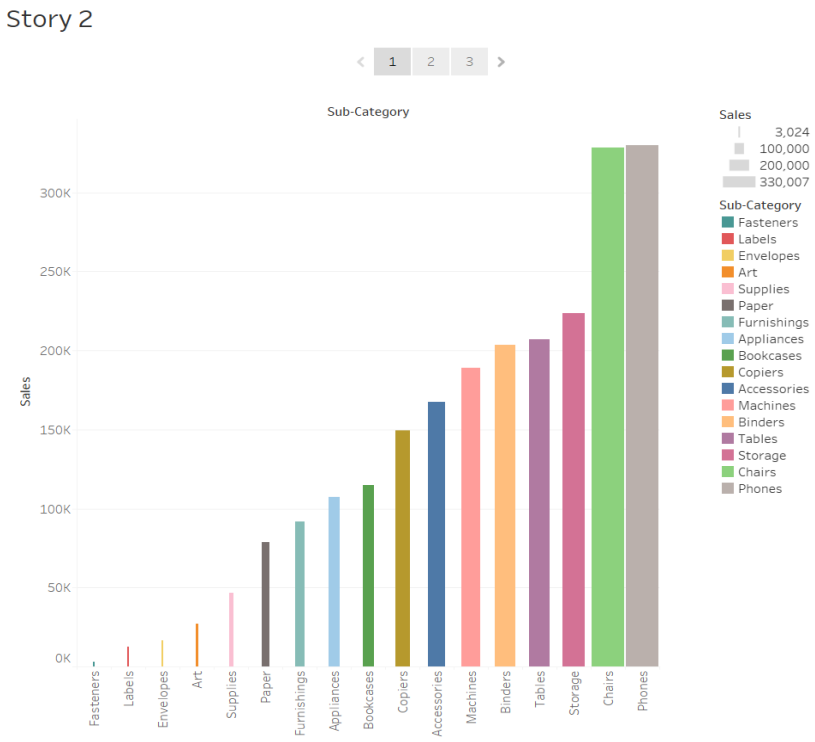
Explanation video link:

https://drive.google.com/file/d/12W1BDNTkyGAE55V8EWtY9ubID7Ms9m6W/view?usp=drivesdk




Story
Layout
New story point
Blank
Duplicate
sheet
line chart
PIE CHART
Sheet 4
WROLD CLOUD
Dashboard 1
Sheet 6
tree map
week of end date
A Drag to add text
Show title
Size
Automatic
Show Me
STORY
1
2
3
4
Week of End Date*
Dec 15
Dec 16
Calculation1
160K
140K
120K
100K
80K
60K
40K
20K
0K
159,167
47,213
47,038
26,816
43,532
34,002
49,048
40,722
48,004
47,743
16,197
15,092
4,149
2,301
1,829
6,416
2,071
801
1,308
2,174
1,786
840
2,082
5,176
1,870
2,143
938
3,253
2,527
5,078
322
1,667
6,583
6,815
4,855
2,809
1,236
845
1 null
3 Jan 16
31 Jan 16
28 Feb 16
27 Mar 16
24 Apr 16
22 May 16
19 Jun 16
17 Jul 16
14 Aug 16
11 Sept 16
9 Oct 16
6 Nov 16
4 Dec 16
1 Jan 17
Week of End Date*
Data Source
sheet
line chart
PIE CHART
Sheet 4
WROLD CLOUD
Dashboard 1
Story 1
Sheet 6
tree map
week of end date

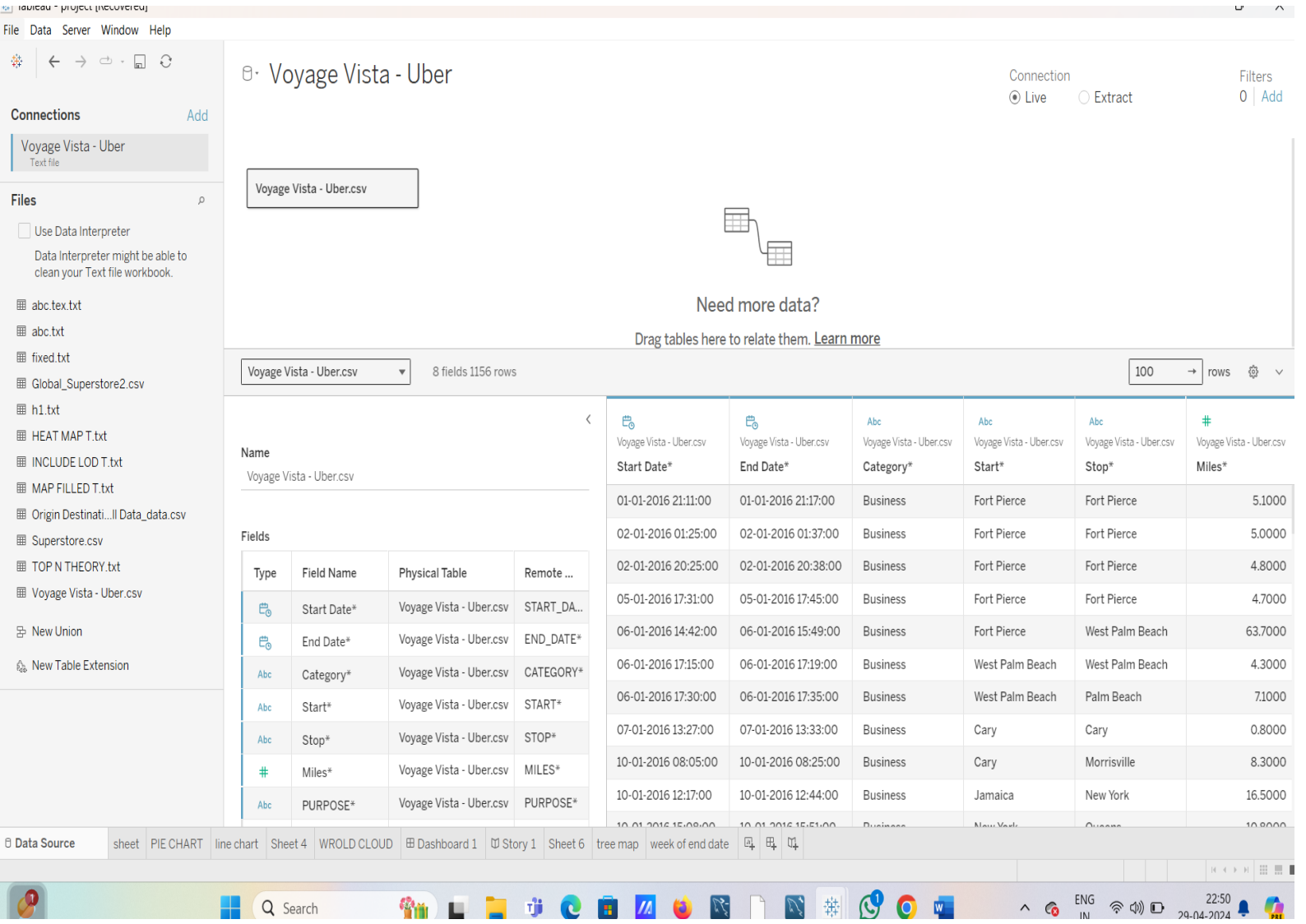
STORY3


Story 2
1
2
3
Sub-Category
Sales
300K
250K
200K
150K
100K
50K
0K
Fasteners
Labels
Envelopes
Art
Supplies
Paper
Furnishings
Appliances
Bookcases
Copiers
Accessories
Machines
Binders
Tables
Storage
Chairs
Phones
Sales
3,024
100,000
200,000
330,007
Sub-Category
Fasteners
Labels
Envelopes
Art
Supplies
Paper
Furnishings
Appliances
Bookcases
Copiers
Accessories
Machines
Binders
Tables
Storage
Chairs
Phones

Milestone 6: Performance Testing

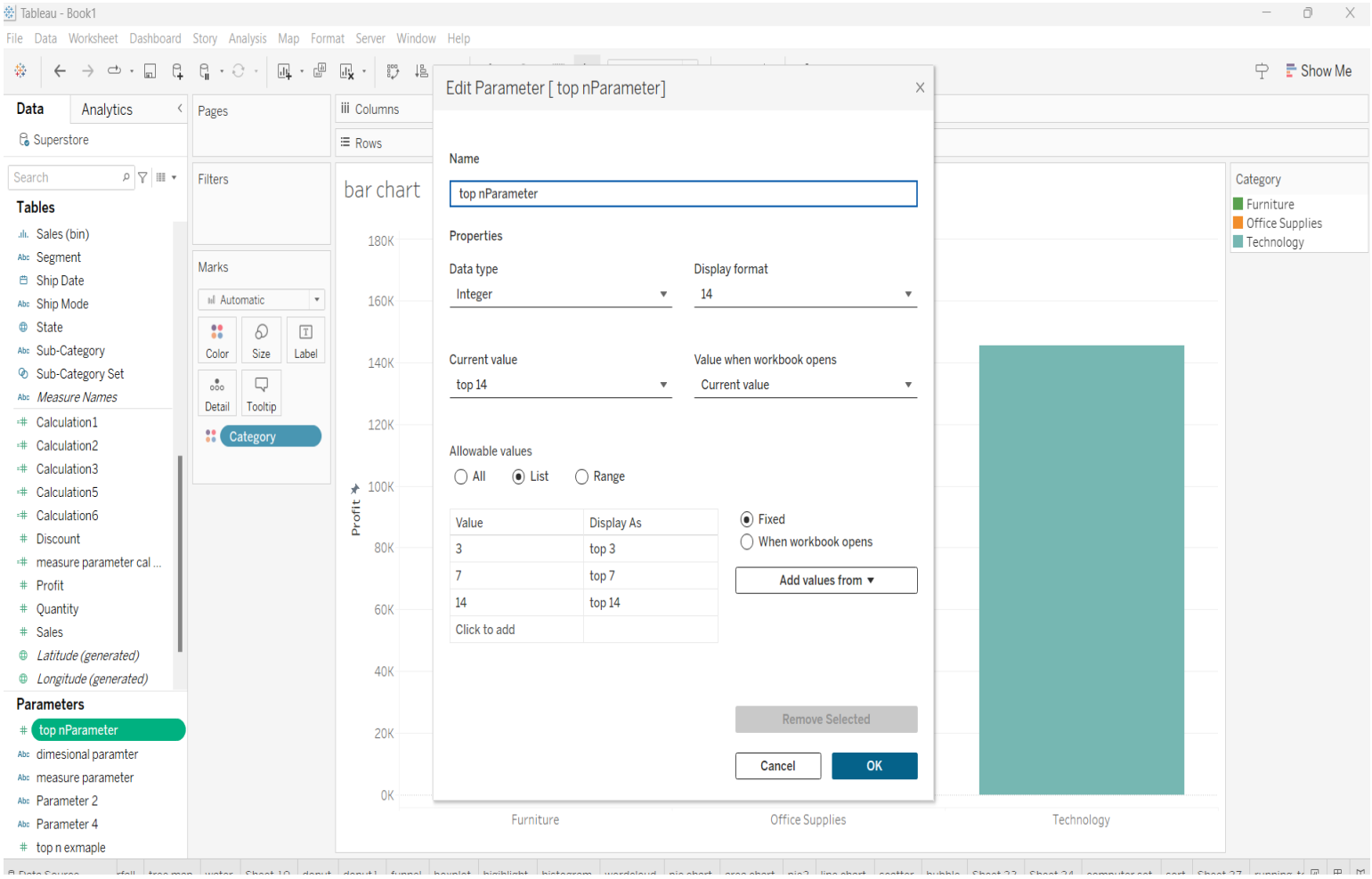
Activity 1: Amount of Data Loaded

"Amount of Data Loaded" refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. It's a measure of how much data has been successfully processed and made available for analysis, manipulation, or use within the system.



Activity 2: No of Calculation Fields

In Tableau, a set is a subset of data based on specific conditions or criteria. Sets allow you to group related data points together for analysis, comparison, or visualization. Sets can be created



dynamically or manually and can be based on dimensions or measures in your dataset.

dynamically or manually and can be based on dimensions or measures in your dataset.

Activity 4: No of Visualizations/ Graphs

1.Number of vayage visit uber

2. Number of –category

3.Total count of the miles for uber.

4.vayage uber of the measure values for uber

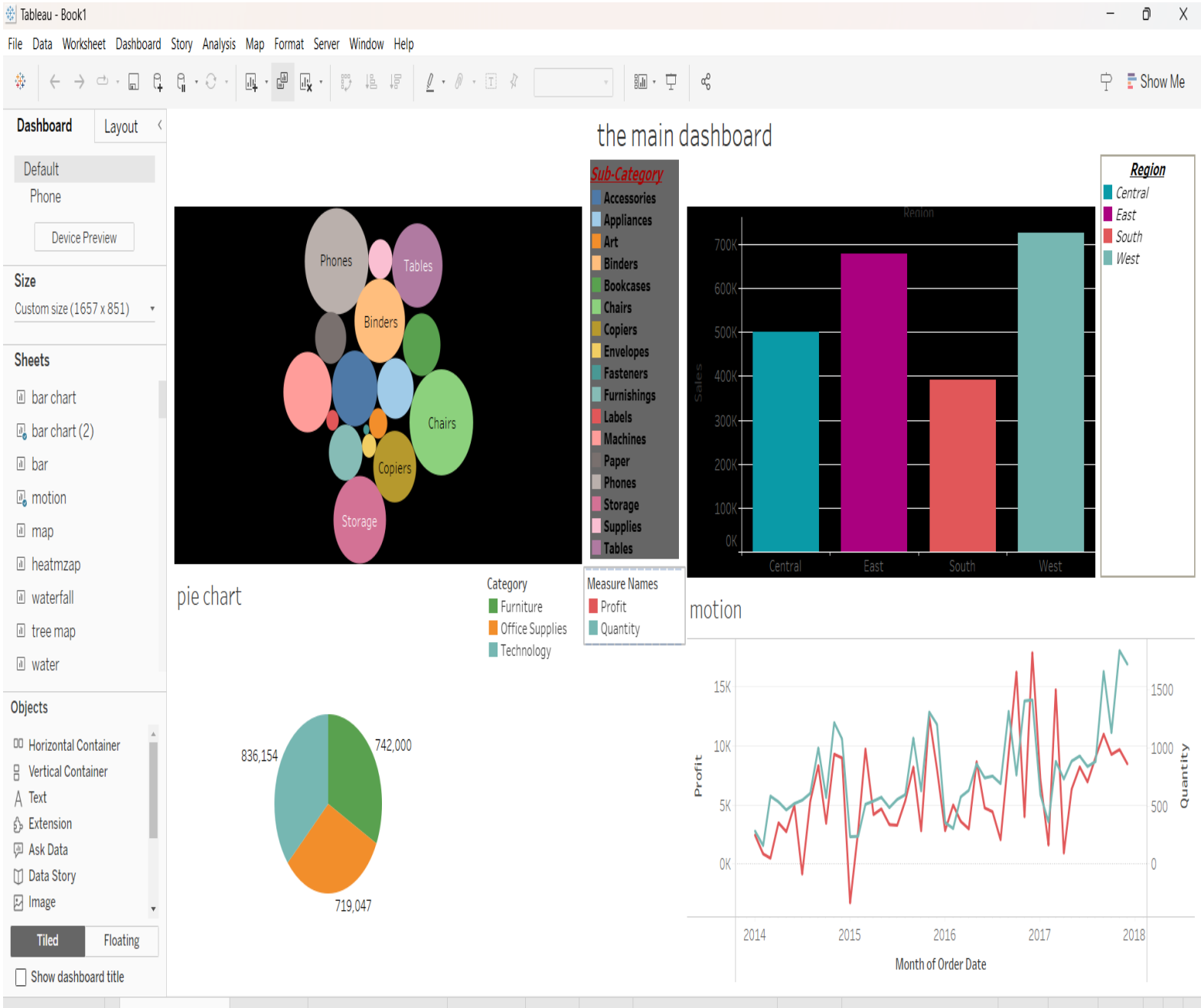
5.Total number starting and ending date

Milestone 7: Web integration

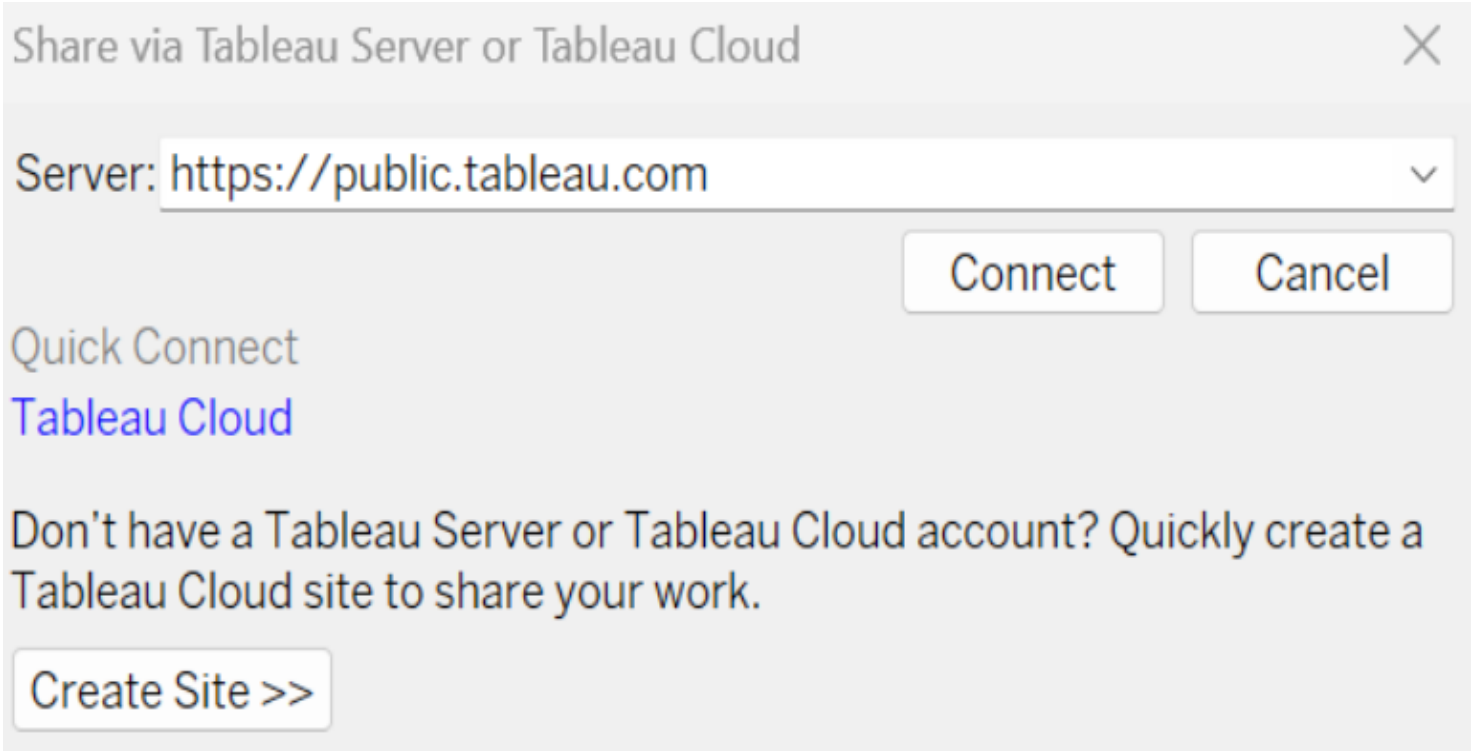
Publishing helps us to track and monitor key performance metrics, to communicate results and progress. help a publisher stay informed, make better decisions, and communicate their performance to others.

Publishing dashboard and reports to tableau public

Step 1: Go to Dashboard/story, click on share button on the top ribbon



STEP 2: LOGIN INTO TABLAEU PUBLIC



Give the server address of your tableau public account and click on connect.

Step 3: Once you click on connect it will ask you for tableau public user name and password

Once you login into your tableau public using the credentials, the particular visualization will be published into tableau public

> **Note: While publishing the visualization to the public, the respective sheet will get published when you click on share option.**

Activity 1: Dashboard and Story embed with UI With Flask

Explanation video link:

https://drive.google.com/file/d/12W1BDNTkyGAE55V8EWtY9ubID7Ms9m6W/view?usp=drivesdk

index.html temp.py

```
1  from flask import Flask, render_template
2
3  app = Flask(__name__)
4
5  @app.route('/')
6  def index():
7      return render_template('index.html')
8
9  if __name__ == '__main__':
0      app.run(debug=True,port=5000)
1
```

Uber Data Analysis

ProjectPro