IBM Cognos To Analyse and Visualize New York City Bike Ride Share Data

A PROJECT REPORT

Submitted by:

- 19BCG10012: Aadira Anil Ramakrishnan
- 19BCE10220: Chirag Sharma
- 19BCE10409: Divya Pratap Singh Bhadoria

LIST OF FIGURES

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
1	TABLE OF CONTENT	4
2	DATA AND ITS FEATURES	5

TABLE OF CONTENT

INTRODUCTION

OVERVIEW

This project is to Analyse and Visualize New York City Bike Ride Share Data. Using the Citi-Bike Analysis dataset, we plan to create a dashboard showing the sales and profits for different segments and Sub-Category of products across all the states.

The data we worked on had the following features:

Trip Duration (seconds)	How long a trip lasted
Start Time and Date	Self-explanatory
Stop Time and Date	Self-explanatory
Start Station Name	Self-explanatory
End Station Name	Self-explanatory
Station ID	Unique identifier for each station
Station Lat/Long	Coordinates
Bike ID	Unique identifier for each bike
User Type (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member)	Customers are usually tourists; subscribers are usually NYC residents
Gender (Zero=unknown; 1=male; 2=female)	Usually unknown for customers since they often sign up at a kiosk
Year of Birth	Self-entered, not validated by an ID.

PURPOSE

To create data visualization charts like those mentioned below:

- Number of trips: Using the Citi-Bike Analysis dataset, we plan to create a dashboard showing the sales and profits for different segments and Sub-Category of products across all the states.
- Customer and Subscriber with Gender: The column User type consists of Subscribers and Customers categories. We plot a stacked bar chart showing the different gender counts of the user type with the trip duration.
- Bike Usage: Plotting a simple bar chart to show the top 10 most used Bike id along with their trip durations
- Age Group Differentiation By Bike: Calculating the number of bikes used by each age group. We only have the birth year in our data. So, we calculate the age of the person by subtracting the birth year from the current year to make visuals easier to interpret. Then show a text table for age group differentiation by bike.
- Top 10 Start Station Names With Respect To Customer Age Group: Finding the top 10 Start station names with respect to the Customer age group and considering the previous analysis we will be plotting a side–by–side bar chart along with the age group.

To create a dashboard using the data visualizations and export the analytics

EXISTING PROBLEMS

- If we are finding unusual patterns within our data analysis or our statistical significance is not strong enough, we might not have enough data to make valid conclusions

- Without doing data analysis, we won't get the opportunity to evaluate the data before making actionable plans

- Data is meaningless without context and without context, we cannot turn data into information

- Information is useless without being able to apply to something
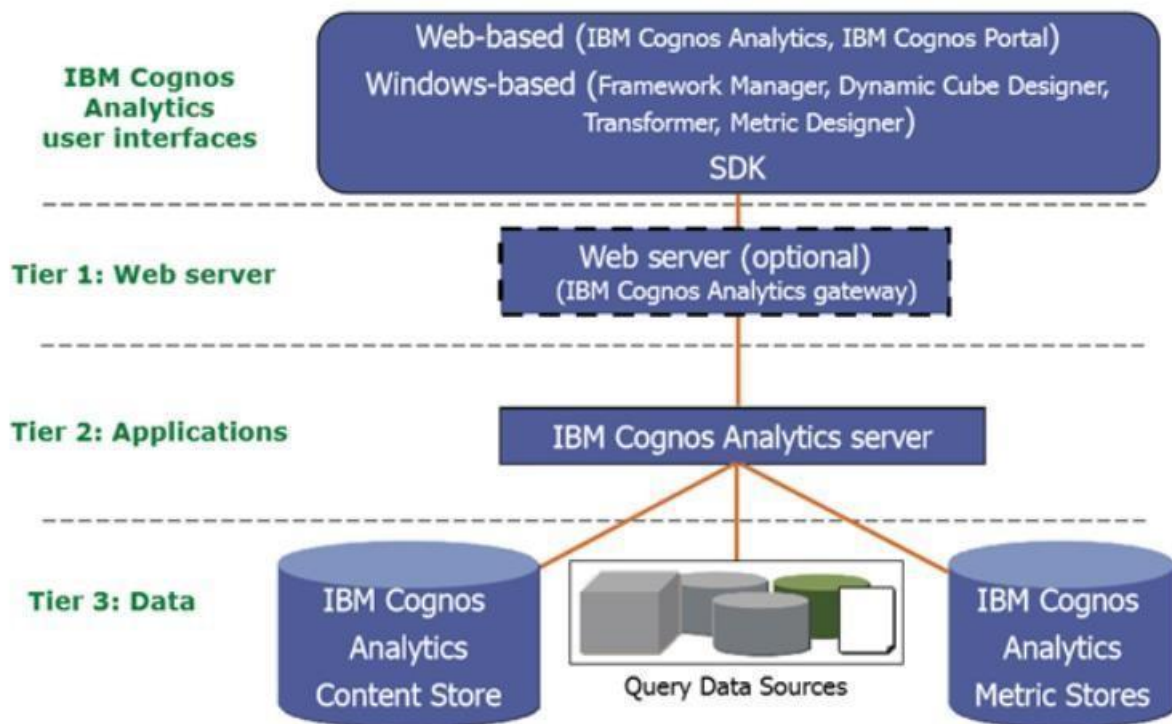
PROPOSED SOLUTIONS

- To create various data visualizations using IBM Cognos

- To make a dashboard using IBM Cognos

- Making dashboards can revolutionize both our success and enjoyment in running our business
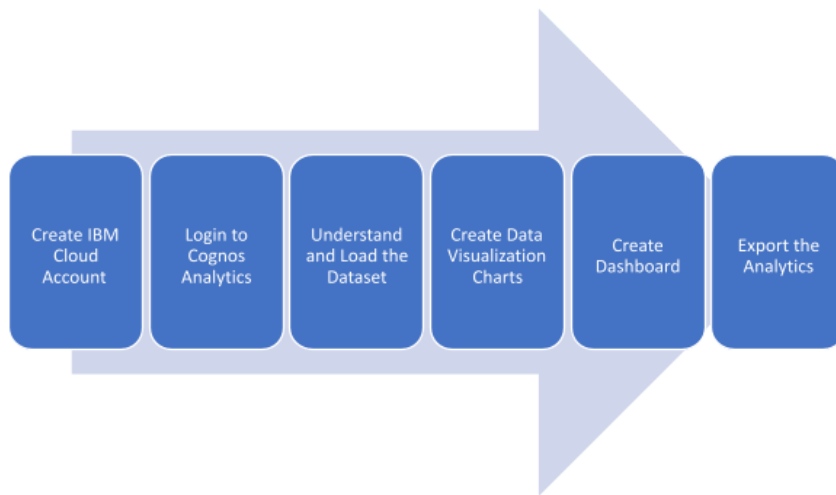
THEORETICAL ANALYSIS

BLOCK DIAGRAM

IBM Cognos Analytics architecture (high level)

IBM Cognos Analytics user interfaces — Web-based (IBM Cognos Analytics, IBM Cognos Portal) Windows-based (Framework Manager, Dynamic Cube Designer, Transformer, Metric Designer) SDK

Tier 1: Web server — Web server (optional) (IBM Cognos Analytics gateway)

Tier 2: Applications — IBM Cognos Analytics server

Tier 3: Data — IBM Cognos Analytics Content Store | Query Data Sources | IBM Cognos Analytics Metric Stores

FLOWCHART

Create IBM Cloud Account → Login to Cognos Analytics → Understand and Load the Dataset → Create Data Visualization Charts → Create Dashboard → Export the Analytics

RESULT

- Number of Trips:

  For visualizing the number of trips, we required the following data.
  - ➔ Bikeid
  - ➔ Trip Duration

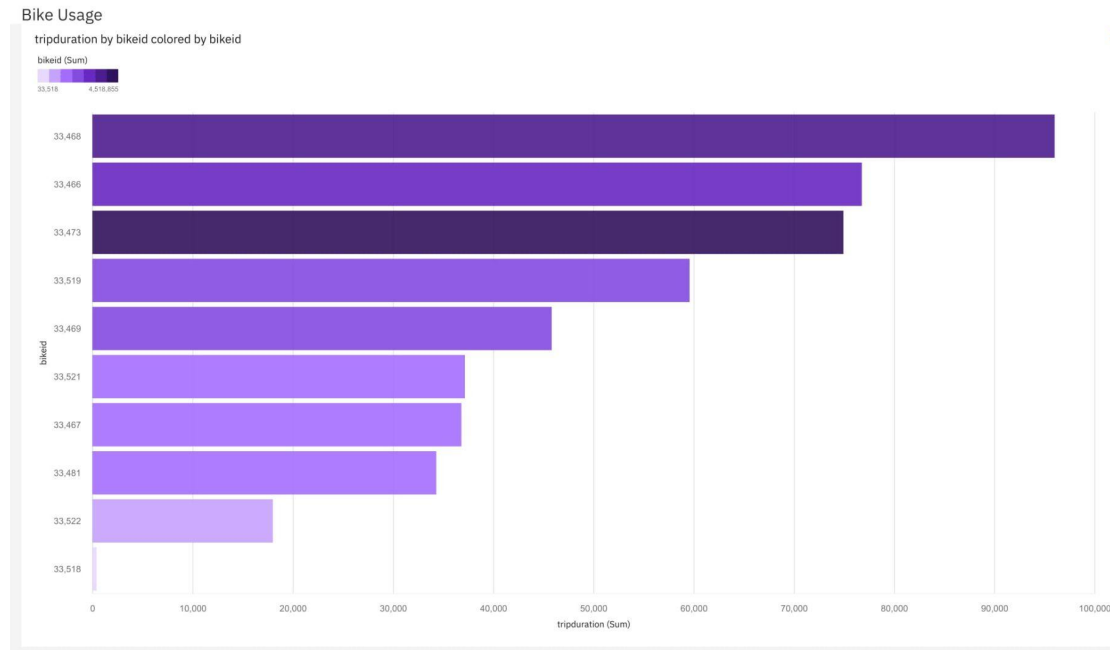 From this data, we plotted a text table visualization.

- Customer And Subscriber With Gender:

  The column User type consists of Subscribers and Customers categories. Here, we plot a stacked bar chart showing the different gender counts of the user type with the trip duration.

● Bike Usage:

We plot a simple bar chart to show the top 10 most used Bike id along with their trip durations.

● Age Group Differentiation By Bike:

We only have the birth year in our data. So, We calculated the age of the person by subtracting the birth year from the current year to make visuals easier to interpret. Then we show a text table for age group differentiation by bike.
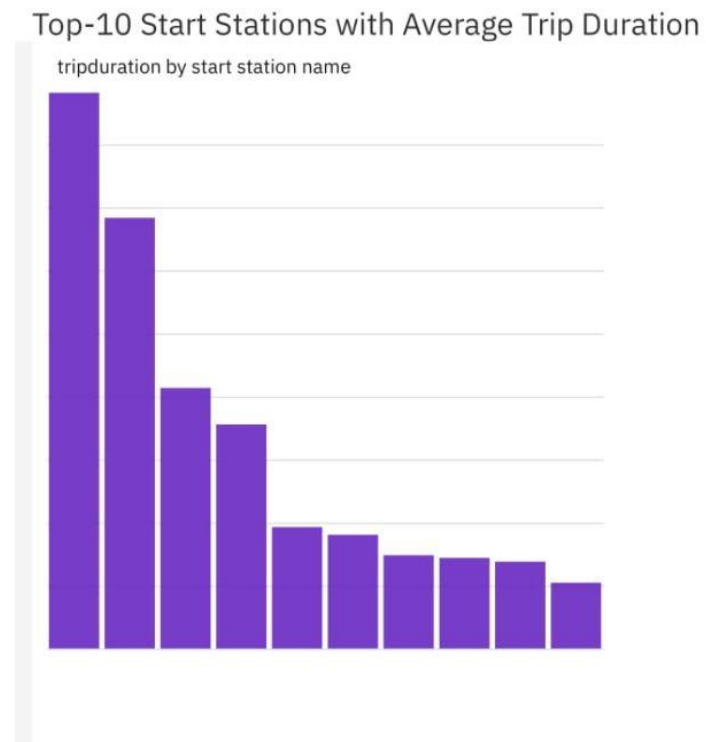
● Top 10 Start Station Names With Respect To Customer Age Group:

We plot a side–by–side bar chart along with the age group.

Top-10 Start Stations with Average Trip Duration

tripduration by start station name

ADVANTAGES AND DISADVANTAGES OF CREATING DASHBOARD

ADVANTAGES

- Enhanced Visibility: Dashboards provide greater visibility with information available whenever it is required to ensure businesses are better placed to respond to changing market conditions
- Timesaving Efficiency: With dashboards, we are no longer wasting valuable time generating reports from multiple systems. Instead, data is drawn from a source and displayed as an easy to interpret visual overview
- Better Forecasting: With greater insight into the data, future demand can be more accurately predicted using historic information. Businesses can be more effectively planned for demand fluctuations, setting measurable goals and deliverables for greater success
- Better Decision Making: Whether you're providing reporting and analysis for the entire organization or functional areas of the business, a dashboard allows companies to analyze key data quickly and meticulously. Visualized interactivity serves to deliver overwhelming amounts of data in a way that is easy to understand. With the ability to easily identify what the data really means; better decisions can be made relevant to the business.

DISADVANTAGES

- Flashy or cluttered design, with users attempting to incorporate too much information without understanding constraints or considering their specific needs from the range of different measurables detailed data analysis provides.
- The technology used in the development of dashboards differs from other software solutions already employed in organizations and can be initially difficult to understand.
- The business has no predetermined rules and hierarchies for how dashboard metrics are used. This means each employee can use the metrics in different ways, resulting in a diverse set of data being reported.

APPLICATIONS

- If you manage complex campaigns, you usually end up having several analytics solutions for each platform and needing to consult them separately, which hinders the overall view. Instead, the dashboard displays data from different sources, like web analytics solutions, social media metrics. This way makes it much easier to compare them and see how they develop.

- A good dashboard clearly shows you many key metrics so you don't need to be an analytics expert to understand them. If you want to look further into a particular data set, you always have the option of employing more specific tools.

- If you synchronize your dashboard automatically in the cloud, you can create different users so that your entire team can access the same information from anywhere. It's even possible to project the dashboard onto a screen in your office so that the whole team can see what is going on in real-time.

- Having a centralized dashboard will save you a lot of time. Instead of collecting data from different sources and making charts on your own, dashboards do all this work for you. You just need to invest some time, in the beginning, to set up the metrics and decide how to present them. From that point on, the reports are created automatically.
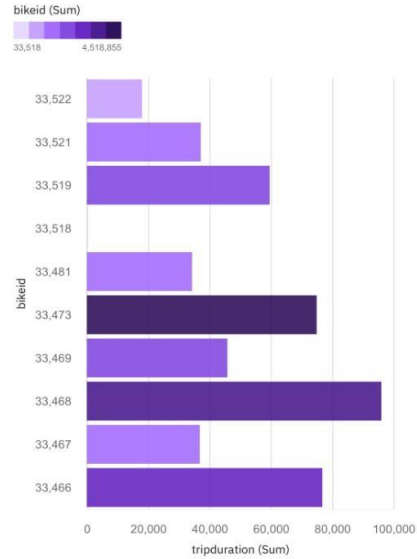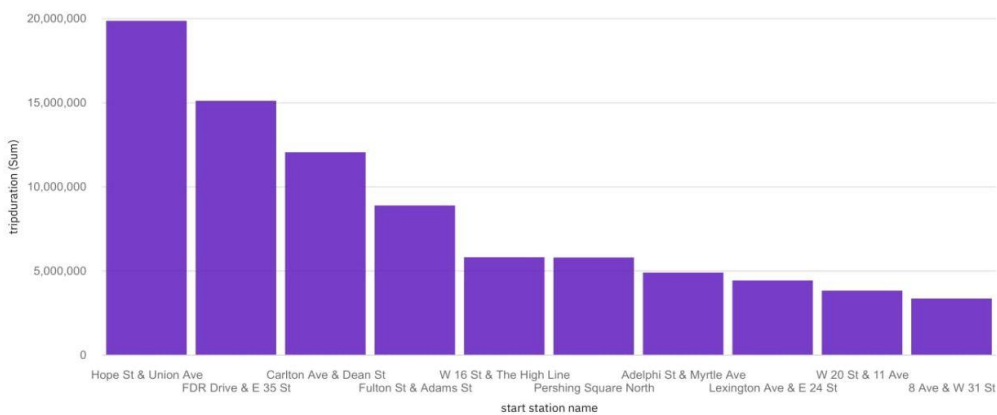
DASHBOARD

Dashboard I

tripduration and bikeid

bikeid	tripduration
10,449	826.81

tripduration by start station name 1

tripduration by bikeid colored by bikeid 2

Age_Group and bikeid

Age_Group	bikeid
21-30	10,190
31-40	10,273
41-55	10,253
<20	4,491
>55	9,764
Summary	10,449

tripduration by usertype colored by usertype

usertype
● Customer ● Subscriber
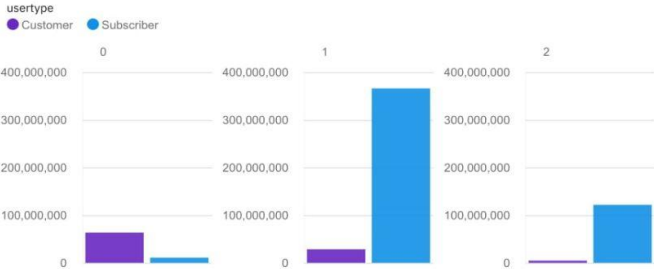
CONCLUSION

This way, with the help of diagrams, graphs, and maps we can understand given data. This understanding of data allows us to ask the right questions to reach our desired goals by optimizing methods. With this project, we learned how to upload and prepare data. We also statistical concepts which helped in calculations and plotting of graphs and maps to make a dashboard.