Bike-sharing in Washington D.C

Geo-data Visualizations

by Mahir Tazwar, Tahmida Saker Muna, Andi Ibrahim

Source: Storyset

Problem Scenario

Source: Storyset

There was company named **Capital Bike Sharing**. This Company wants to expand their business and increase the number of Cycles to approximately **5000** bikes.

Problem Scenario

Source: Storyset

While discussing it on a meeting with business development team, the challenge arrives, how and where would they distribute these extra bikes?

Problem Scenario

Source: Storyset

Possible questions:

High Bike Usage Location

1. Which areas or stations has the higher number of bike usage in terms of bike arrival and departure?

Peak Times/Seasons for Bike Usage

2. In which time of the day, people use bikes the most?

3. Is people's bike usage varying in season?

Datasets

Available Dataset

Raw Dataset

ride_id	rideable_type	started_at	ended_at	start_stationname	start_stationid	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
CA46A1E7C19A98DD	classic_bike	DD/MM/YYYY Hour:Min:Sec	DD/MM/YYYY Hour:Min:Sec	10th & Florida Ave NW	31120	Thomas Circle	31241	38.92039	-77.0257	38.9059	-77.0325	member
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
3F7C1AC5CF515183	classic_bike	########	########	10th & Florida Ave NW	31120	7th & T St NW	31109	38.92039	-77.0257	38.9155	-77.0222	member

- Acquired from *https://capitalbikeshare.com/*
- Data Ranging from 2010 to Present
- For Washington DC and nearby regions
- More than 700 Stations
- More than 6927165 rows
- More than 3.00 GB

Data Preprocessing

Raw Dataset

ride_id	rideable_type	started_at	ended_at	start_stationname	start_stationid	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
CA46A1E7C19A98DD	classic_bike	DD/MM/YYYY Hour:Min:Sec	DD/MM/YYYY Hour:Min:Sec	10th & Florida Ave NW	31120	Thomas Circle	31241	38.92039	-77.0257	38.9059	-77.0325	member
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
3F7C1AC5CF515183	classic_bike	########	########	10th & Florida Ave NW	31120	7th & T St NW	31109	38.92039	-77.0257	38.9155	-77.0222	member

Our Approach

- Only Considered points **within Washington DC boundary**

- Only 359 Station Points

- Considered **only last 2 Years Data**

- July 2021 to July 2023

Data Preprocessing

Raw Dataset

ride_id	rideable_type	started_at	ended_at	start_stationname	start_stationid	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
CA46A1E7C19A98DD	classic_bike	DD/MM/YYYY Hour:Min:Sec	DD/MM/YYYY Hour:Min:Sec	10th & Florida Ave NW	31120	Thomas Circle	31241	38.92039	-77.0257	38.9059	-77.0325	member
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
3F7C1AC5CF515183	classic_bike	########	########	10th & Florida Ave NW	31120	7th & T St NW	31109	38.92039	-77.0257	38.9155	-77.0222	member

Start Time

Seasons

- Spring ---------- March to May
- Summer ---------- June to August
- Fall ---------- Sept to Nov
- Winter ---------- Dec to Feb

Type of Weekday

- Weekday
- Weekend

Hours

- 0 to 23

Data Preprocessing

Raw Dataset

ride_id	rideable_type	started_at	ended_at	start_stationname	start_stationid	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual
CA46A1E7C19A98DD	classic_bike	DD/MM/YYYY Hour:Min:Sec	DD/MM/YYYY Hour:Min:Sec	10th & Florida Ave NW	31120	Thomas Circle	31241	38.92039	-77.0257	38.9059	-77.0325	member
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
3F7C1AC5CF515183	classic_bike	########	########	10th & Florida Ave NW	31120	7th & T St NW	31109	38.92039	-77.0257	38.9155	-77.0222	member

Tools Used: Python, Pandas, Excel

Station ID	Departure Count	Arrival Count	Seasons	Day Type	...	Ward
Station 1	56	07	Winter	Weekday	...	Ward 2
					...	...
Station 359	105	68	Summer	Weekend	...	Ward 4

Processed Dataset

Software Choice

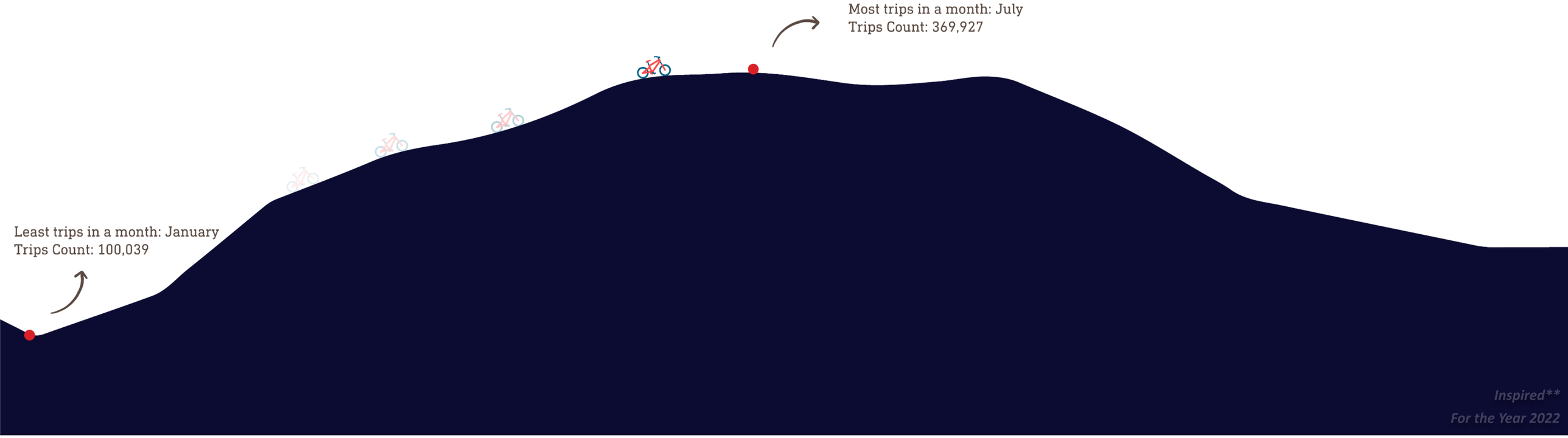
- It can connect easily to spreadsheets and handle large datasets.
- Customization and flexibility, allowing users to tailor visualizations.
- Easy sharing online to overview the visualizations and dashboard within teams.
- Ensuring reproducibility

Exporation of Bike Station wise User Dynamics from 2021 to 2023

Capital BikeShare: WashingtonDC

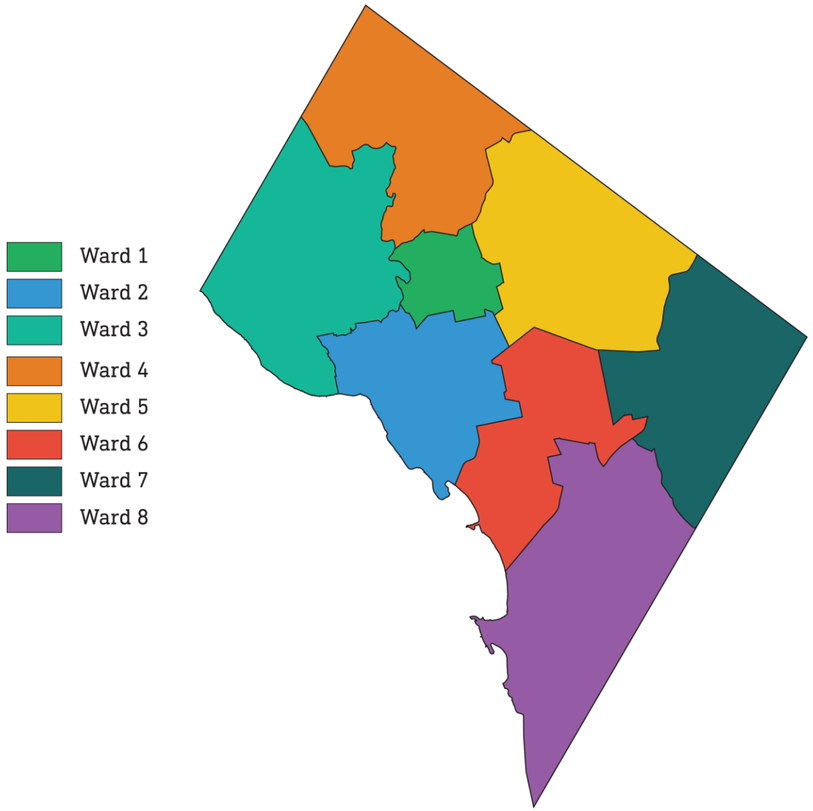
OVERVIEW

DETAILED



Solution Overview

Wards of Washington DC



***Visualization Tool**: ArcGIS Pro*

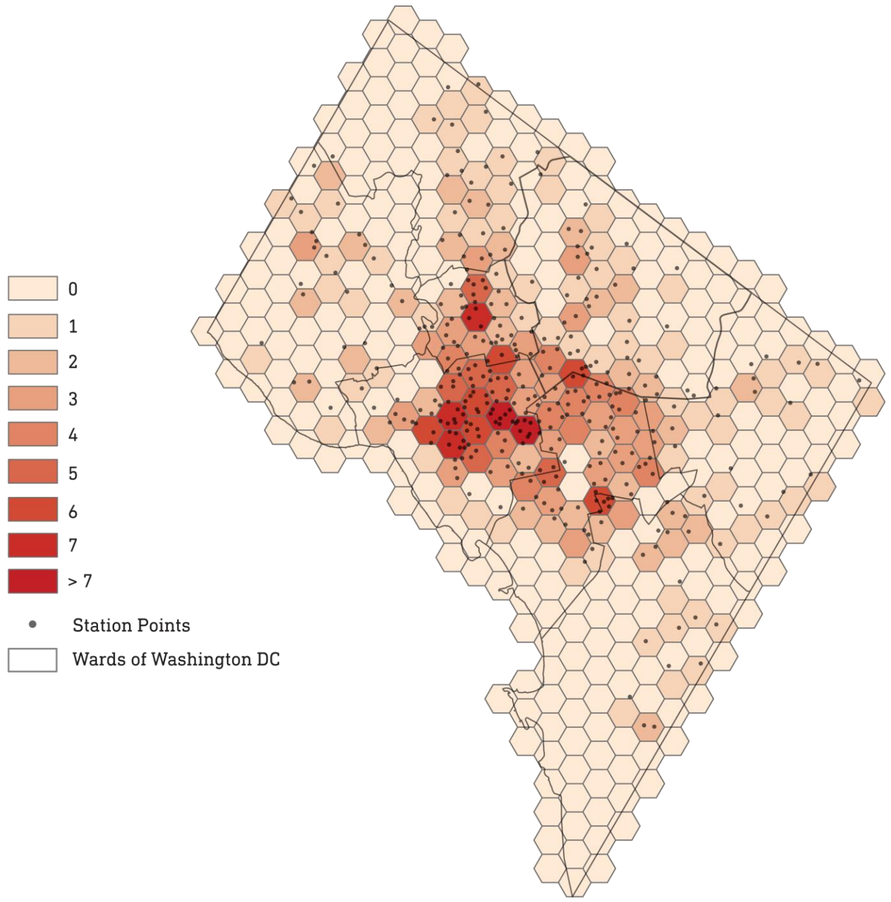
Considerations

- Choropleth Map – Different Color denotes Different value (different wards)

Data Source : OpenData DC

Solution Overview

Station Distribution Map



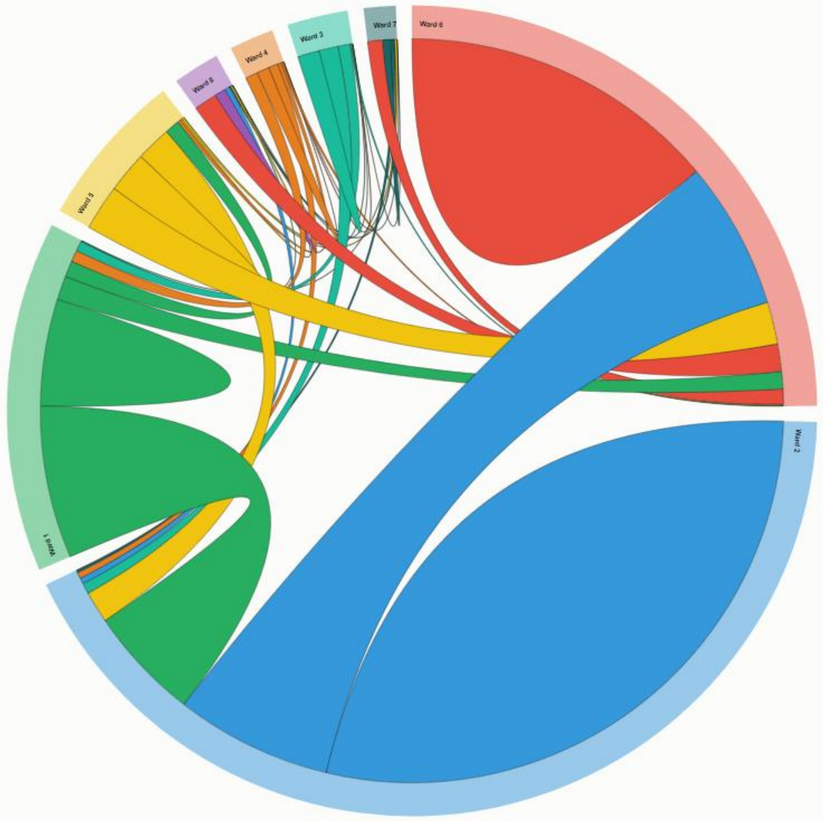
***Visualization Tool**: ArcGIS Pro*

Considerations

- Hexagon Dot Density Map - One Hexagon represents 0.5 Kilometer distance
- Continuous color scheme to project different density
- Equal Area Projection for showing the distance

Solution Overview

Chord Diagram Showing the Frequency of Trips by Wards



***Visualization Tool**: D3 (with Help of ChatGPT and Observable)*

Based on Start and End Stations, we organized the data into such manner

Example

	Ward 4	Ward 8	Ward 3	Ward 5
Ward 4	50	15	100	25
Ward 8	25	5	30	2
Ward 3	60	65	7	19
Ward 5	15	14	29	35

Considerations

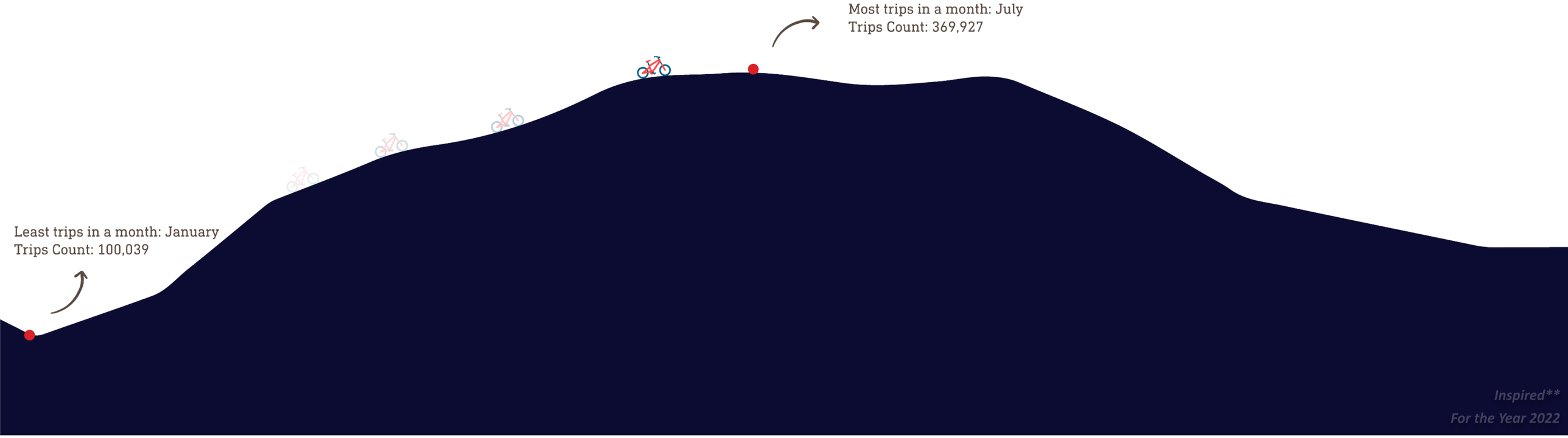
- Same colors for same Ward – Easier for users to relate
- Interaction able interface to observe the values

Exporation of Bike Station wise User Dynamics from 2021 to 2023

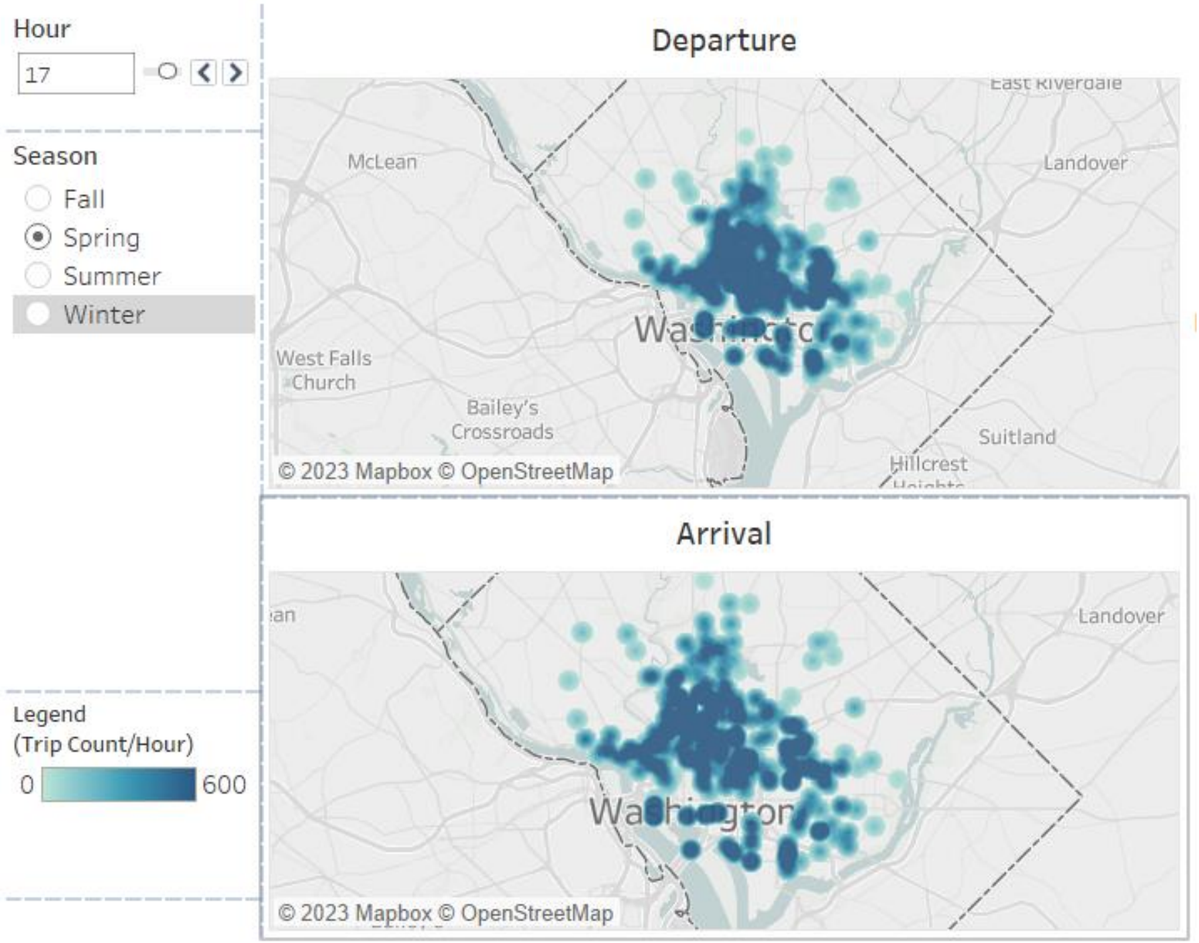
Capital BikeShare: WashingtonDC

OVERVIEW

DETAILED



Detailed Solution



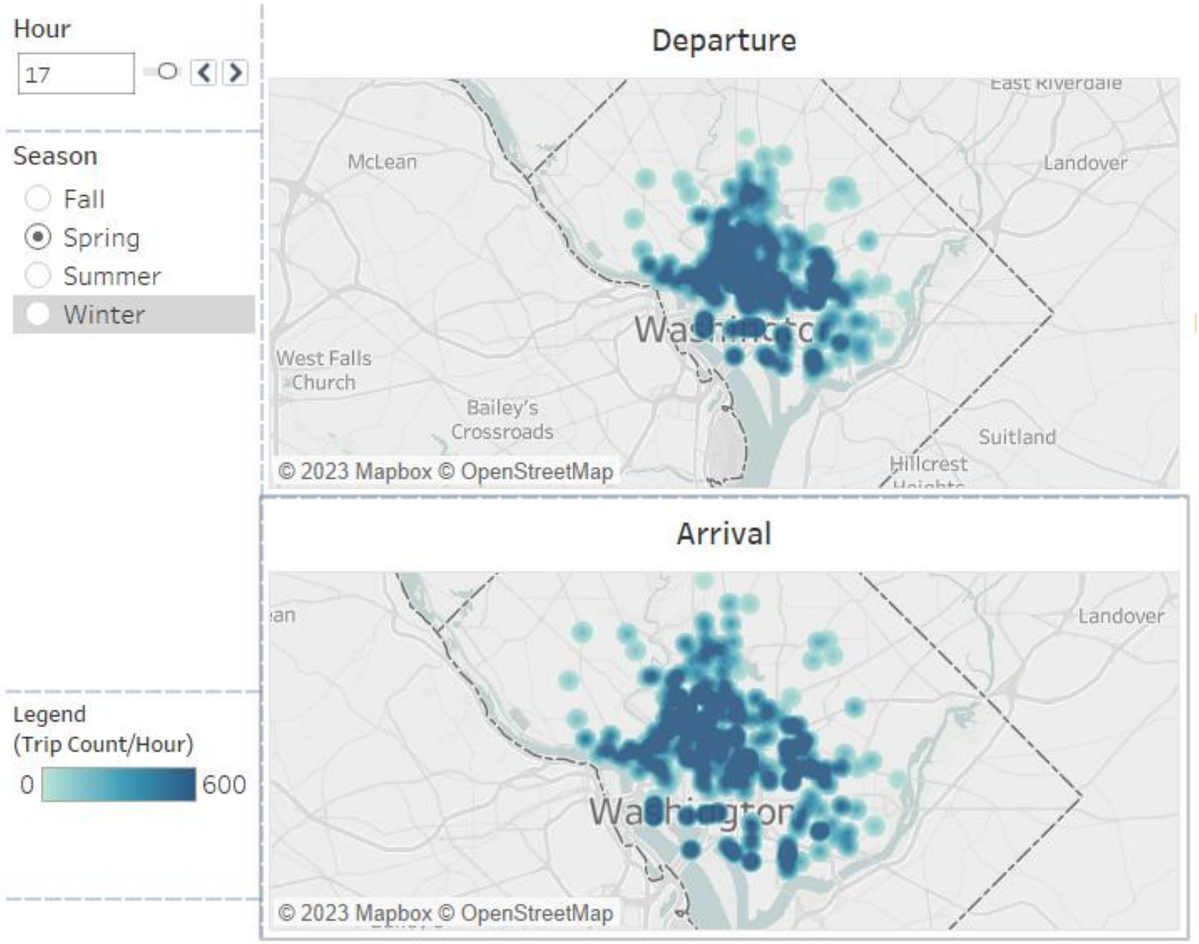
Visualization Tool: Tableau

Visualization 1: Density Map

Consideration

- Value Visualization
  - Density/Heat Map – Shows Concentration of data and effective alternative of point visualization
  - Continuous Sequential color Scheme (Blue Teal) for showing the high-density and low-density value
- Background Map
  - Street
  - Cities Name
  - Area Code Labels
  - State Boundary

Detailed Solution



Visualization Tool: Tableau

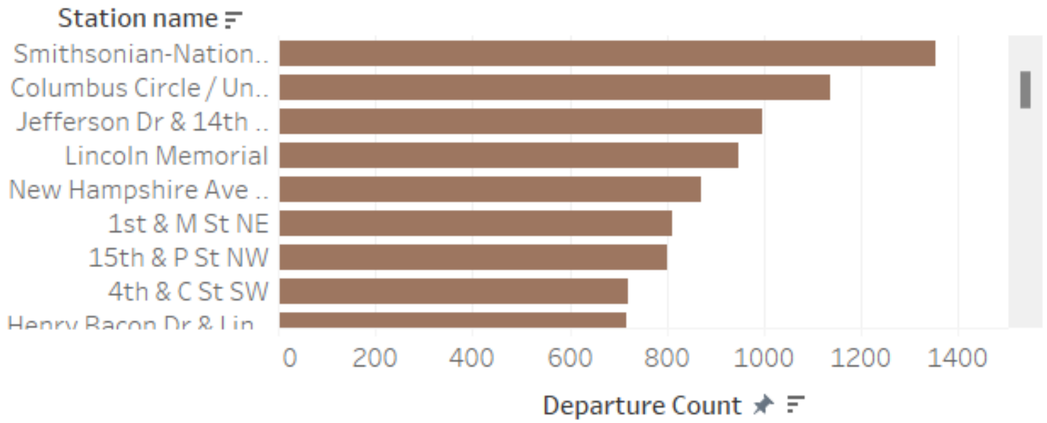
Visualization 1: Density Map

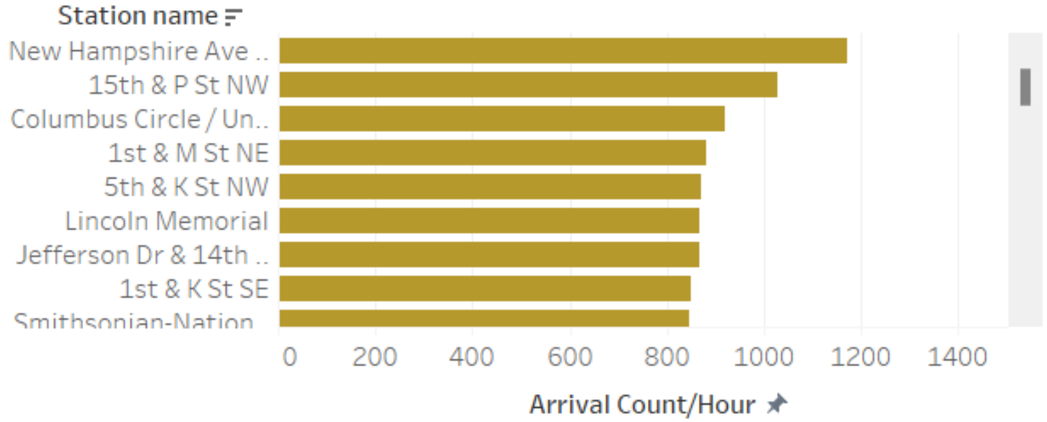
Reflection

- Initially we used the **visual variable size** to show to Arrival and Departure value. But Size doesn't represent the variation in value.
- We had also a challenge of **different legend in different station** which give us a varying idea of high and low density.

Same color scheme for easy comparison.

Detailed Solution





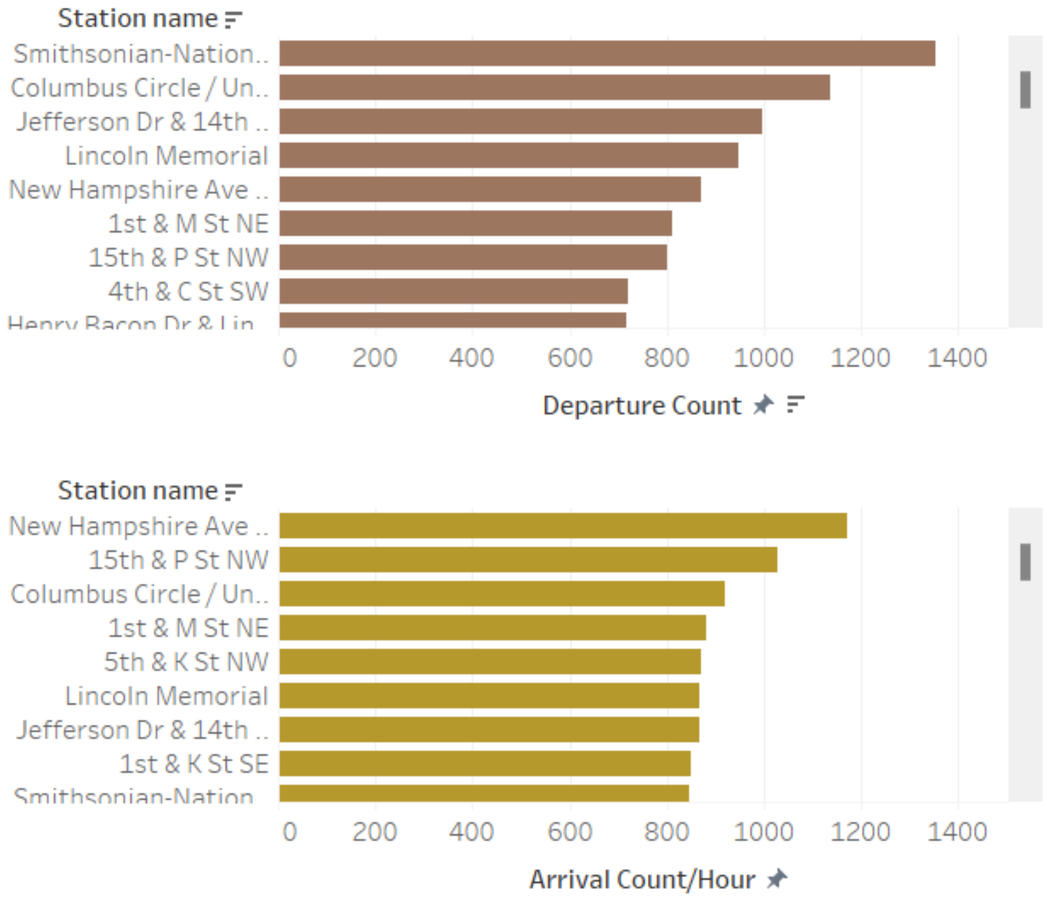
Visualization Tool: Tableau

Visualization 2: Bar Chart for Arrival/Departure Count (Each station)

Consideration

- Horizontal Bar chart – can accommodate all the station horizontally, doesn't make any cluttering
- The station names are easily readable
- Data is sorted from highest to lowest for both cases

Detailed Solution



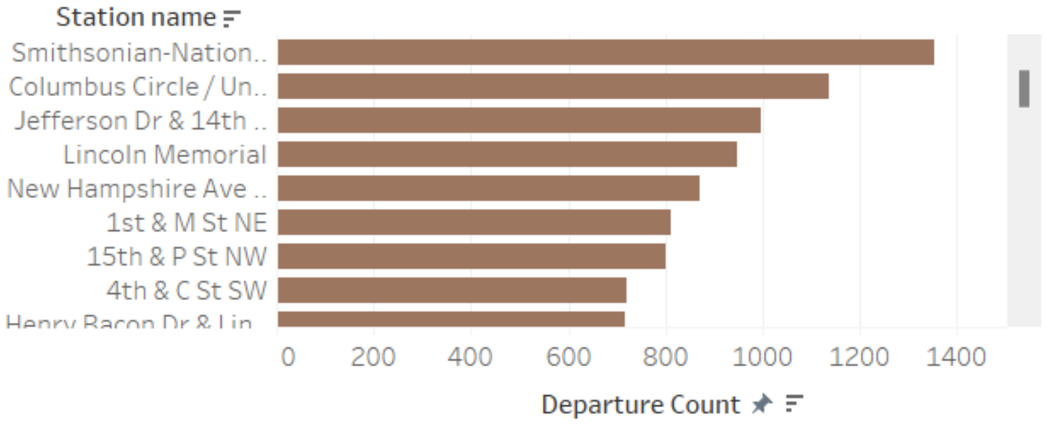
Visualization Tool: Tableau

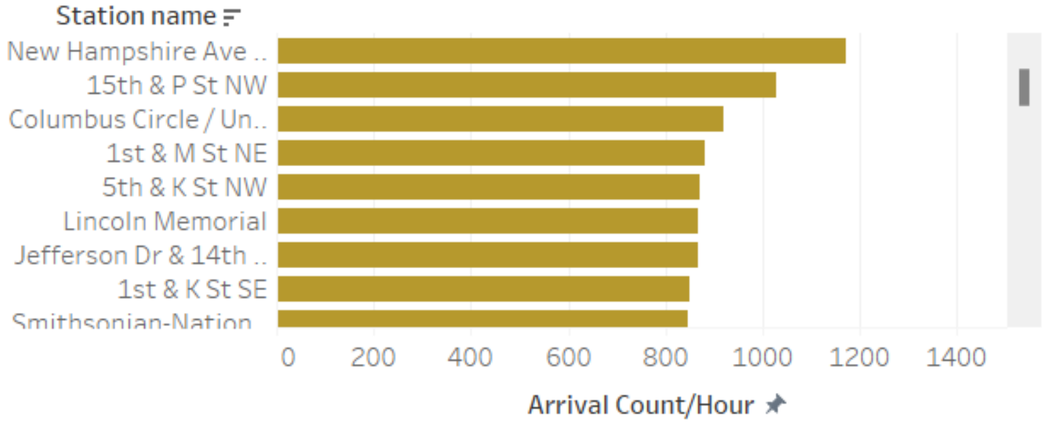
Visualization 2: Bar Chart for Arrival/Departure Count (Each station)

Consideration

- The X-axis has same value range - starts from zero and ends at 1400
- Single colored bar chart is used for both Departure and arrival
- A filter is given to check the value for each location one by one for each hour

Detailed Solution





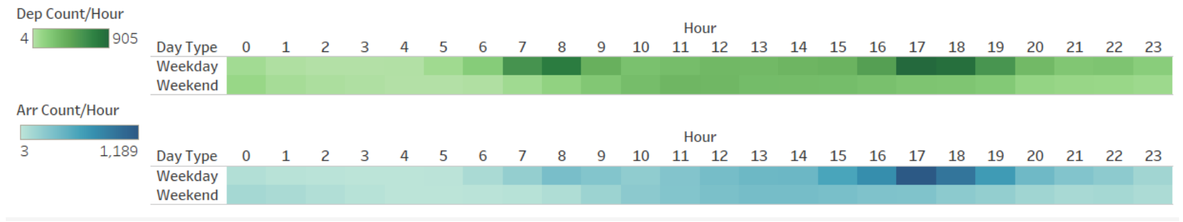
Visualization Tool: Tableau

Visualization 2: Bar Chart for Arrival/Departure Count (Each station)

Reflection

- Here, we had also a challenge of **different axis range in different station** which give us a varying idea of high and low count.
- Initially we made a **relation between each station of the Density map and each station of the Bar chart**. But the gradient color in each point might confuse our user.

Detailed Solution



Visualization Tool: Tableau

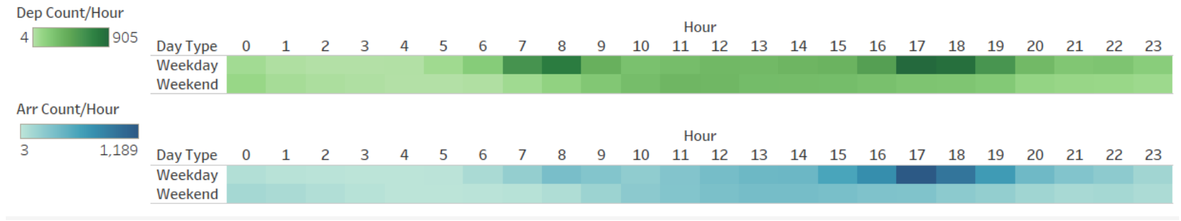
Visualization 3: Heat Map

Consideration

- Sequential Color Palette is used as the data has no meaningful threshold
- All hours are evenly distributed and will help our user to see the change one station at a time
- Different legend range has been used for better portraying the dynamics in weak

Detailed Solution

Visualization 3: Heat Map



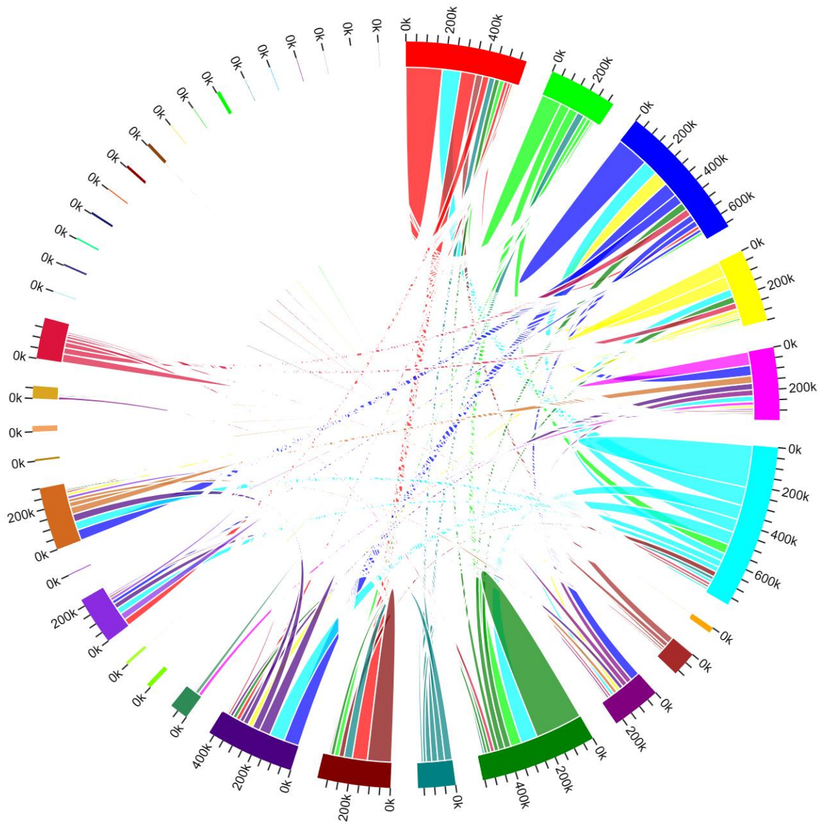
Visualization Tool: Tableau

Reflection

- A potential problem of this heat map is that it shows nearly **similar color for close values**.
- But our datasets have only wide range of positive values, there is **no meaningful threshold**, based on which we can make divergent color palette.
- Another **trade-off – Different legend for different station** for better understanding the dynamics

Solution Overview

Chord Diagram Showing the Frequency of Trips by Neighborhoods



***Visualization Tool**: D3 (with Help of ChatGPT and Observable)*

Final Reflection

Dashboard

Final Reflection

Do our visualizations provide answers to the questions we have?

Inspirations

- https://observablehq.com/@d3/gallery?utm_source=d3jsorg&utm_medium=nav&utm_campaign=try-observable
- https://bost.ocks.org/mike/uberdata/
- https://www.reddit.com/r/tableau/
- https://www.youtube.com/watch?v=cMNcFBagc9M&ab_channel=DataScienceTutorials

Thank You