Section 6, Group 4

Boston Red Sox Home Games & Demand for Rental Bicycles

By: William Bonner, Ke-Chu Chen, Truong Dinh, Erica Elias, Saif Ali Khan, Katie Schlafhauser

Section 6, Group 4

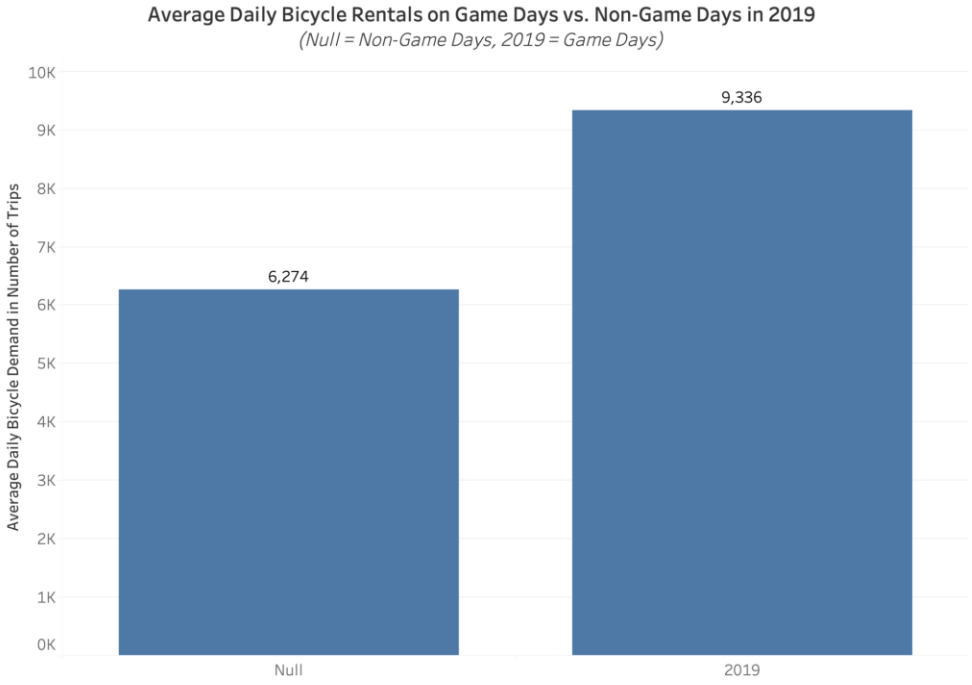
Introduction

Fenway Park is the oldest Major League Baseball (MLB) stadium still in use by a professional team, making its first debut in 1912. The ballpark and its home team, the Red Sox, continue to draw in large crowds from near and far each game, requiring various forms of transportation to allow fans to watch the game and see the “Green Monster”, a 37-foot tall wall in left field. As a group, we were interested in learning more about the relationship between Red Sox home games and the demand for bicycle rentals in the Boston area.

The prominent business questions that arose for Blue Bikes were: 1) Is Blue Bikes optimizing its operation to capture traffic moving to and around the stadium during Red Sox home games? 2) Is Blue Bikes marketing to the right demographics or through the appropriate channels? We developed several hypotheses and conducted analysis to test these hypotheses. The hypotheses are as follows:

1. Bicycle demand will be higher on days with a Red Sox home game than days without a home game
2. Bicycle demand will be stronger in areas immediately around the stadium on game days vs. non-game days
3. A higher proportion of males will rent on game days than females
4. Younger individuals will rent at higher rates on game days than older people

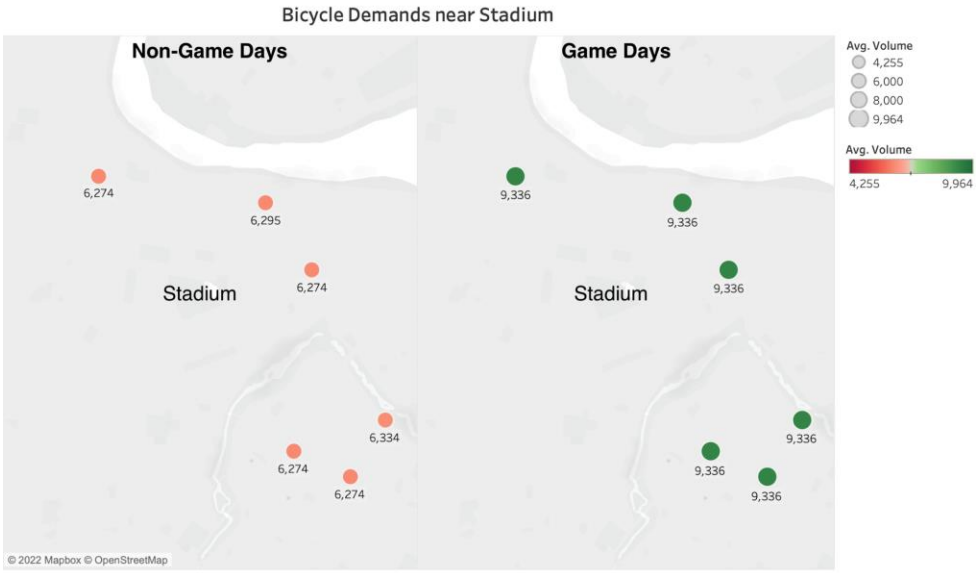
Hypothesis 1: Bicycle demand will be higher on days with a Red Sox home game than days without a home game.



Takeaways: On average, there are a higher number of bicycle trip rentals on game days versus non-game days (~49% more). We believe this may be due to a higher number of visitors entering the city on game days that may do more activities in addition to going to the game *(e.g., going out to dinner, tailgating and renting a bike to get to the game, touring Boston, etc.)*.

However, despite this clear increased demand, it is important to note that game days primarily happen during the summer (June, July, & August) when there may be higher numbers of visitors or bike users in Boston anyways. The non-game day data looks at every date in 2019, and may under-represent the true average daily rentals during a comparable time of the year.
Business Recommendation: In order to accommodate increased demand on game days, ensure operations teams are optimizing the placement and replacement of bicycles on days in which there is a home game.
Data Sources: To create this visual, we pulled data from the SportsRadar API, and two separate SQL data tables from the Blue Bikes Boston Data. We linked all three files together by date, and filtered by the sports game schedule. Any of the dates on which there was a baseball game, the aggregate bike rental data showed up under “2019” since it could be attributed to a found date in the baseball schedule. We attempted to adjust the labeling in Tableau, but it wouldn’t allow us to edit the Alias of a date function. In our Tableau visual, to calculate the average number of trips per day, we downloaded a consolidated file from SQL which aggregated the number of unique trips per day. We then added in the date from the scheduled baseball games data and used the dates as a horizontal axis (as noted above - null = any non-game day and 2019 = any game day).

Hypothesis 2: Bicycle demand will be stronger in areas immediately around the stadium on game days vs. non-game days.

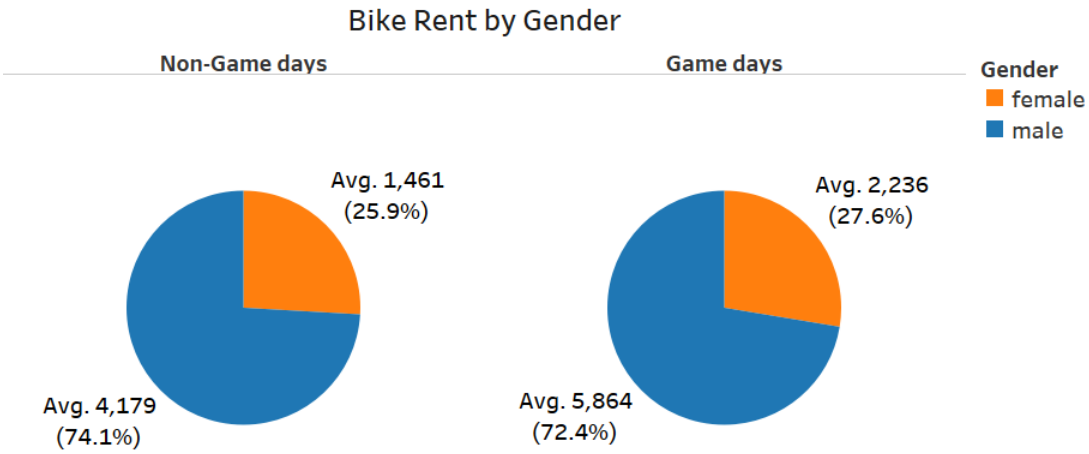


Takeaways: On average, there was a higher number of bicycle trip demand near the stadium on game days versus non-game days (~48% more) in 2019. This may be because people are more likely to rent bicycles on game days to tailgate and gather in and around the stadium on game days, in addition to making their way directly to the stadium. There is a possibility that the higher demand may also be caused by higher numbers of tourists in Boston or general bike users, since the season primarily occurs in the summer, when seasonally, bicycle stations have higher demand.

Business Recommendation: To accommodate high dropoff rates of bicycles at stations near the stadium, be sure operations employees are prepared to quickly transition the bicycles to other stations of high demand, so that users have spaces to drop off their bicycles and so that bicycles at other stations are available for rent for customers wanting to cycle to the stadium.

Data Sources:
We used the data from the SportsRadar API, and two separate SQL data tables from the Blue Bikes Boston Data. The information needed from Bluebikes includes Dates, the number of rent on different dates, latitude and longitude. Please see the SQL code in the Appendix (Hypothesis 2 Appendix Items) to process the Bluebikes database.

Hypothesis 3: A higher proportion of males will rent on game days than females.

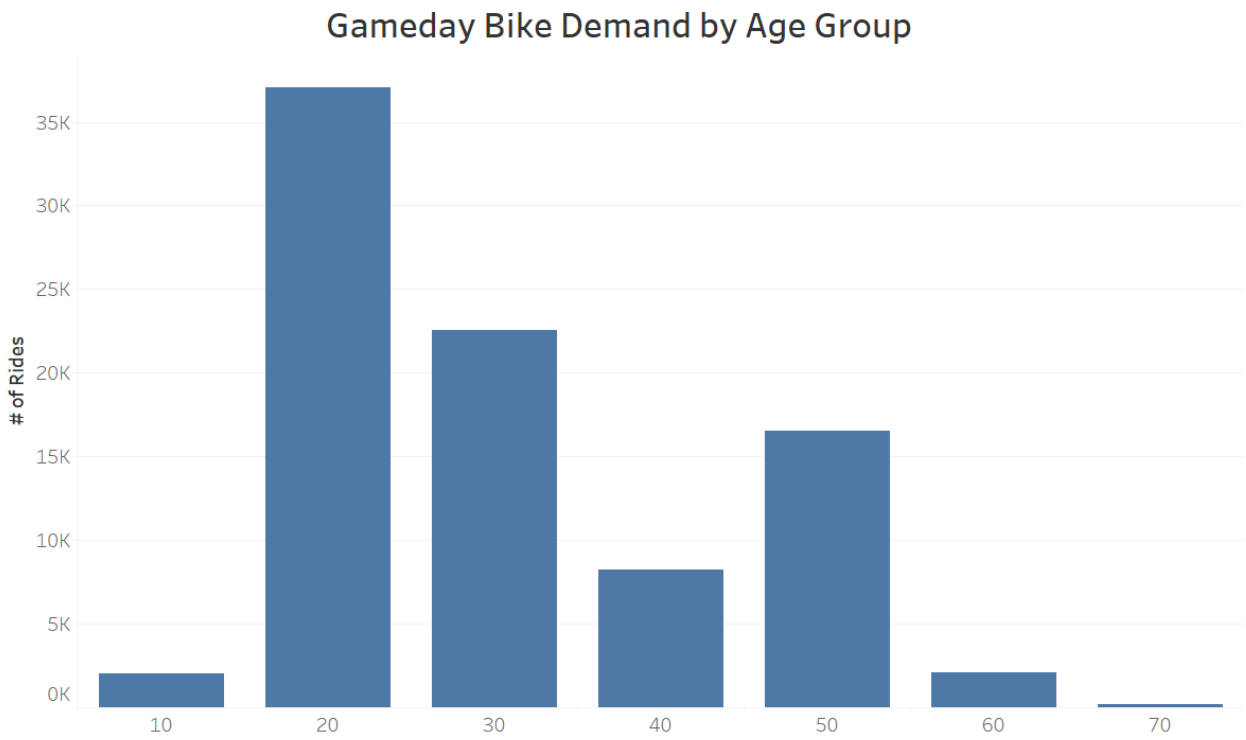


Takeaways: Males account for the majority of bike rental. On game days, the average number of bikes rented increased for both genders. Additionally, the proportion of female renters increased by nearly 2% (from 25.9% to 27.6% of renters).

Business Recommendation: Expand upon female-oriented marketing strategies to capture additional customers.

Data sources: The chart was created by combining processed data from the Blue Bikes database and Game day data from Sportradar API (appendix). The information needed from Bluebikes includes dates, gender, and the number of rent by date and gender. Please see the SQL code in the Appendix (Hypothesis 3 Appendix Items) to process the Bluebikes database.

Hypothesis 4: Younger individuals will rent at higher rates on game days than older people.



Takeaway: The majority of gameday bike demand is from customers 20-29 years old. This is in line with our hypothesis that younger individuals will rent at a higher rate on game days compared to older people. However, it's interesting to see customers in their 50s have higher gameday bike demand compared to customers in their 40s. We believe that this can be attributed to more owners in their 40s having families with young children and requiring a vehicle to transport the young family to the game. Customers in their 50s are closer to retirement and those with children have already moved out of the house. More customers in this segment do not require a vehicle to transport young family members to the game, as compared to customers in their 40s.

Business Recommendation: Blue Bike should consider rolling out a campaign targeted at young families in their 40s. This campaign could include bikes with child seats and storage to entice parents to utilize bikes even when they are traveling with their children. As shown in the Hypothesis 1 Appendix Items, this campaign is relevant on both game days and non-game days (approximately 10% of total demand from customers in their 40s).

Data sources: The chart was created by joining processed data from the Bluebikes station database and game day data from Sportradar API (appendix). The information needed from Bluebikes includes birth year and the count of rentals. A calculated field was created within Tableau to generate the age at time of rental based on the birth year, which was then put into bins with intervals of 10 to easily digest the graph. The information needed from the Sportradar API includes date and status. The date field was used to analyze 2019 data only, while the status field was used to filter for only game days.**Appendix**

Section 6, Group 4

SQL & Python Code Appendix

SQL Data Source: For our project we chose the bluebikes database provided on the phpMyAdmin server (https://phpmyadmin.it.umich.edu/index.php?server=33). We used the following SQL queries in order to generate the data of interest for us: -

1. Join the two tables from bluebikes database based on end station ID and filter for the year 2019:

```
SELECT *, DATE(tr.starttime)
FROM station st INNER JOIN trip tr
ON st.stationId = tr.endStationId
WHERE YEAR(tr.starttime) = 2019;
```

This query joins the two given tables and gives out data of all bike rentals in the year 2019.

2. Group trip data by date:

```
SELECT DATE(starttime) AS rentdate, YEAR(starttime) AS rentyear,count(*) as volume
FROM trip
WHERE YEAR(starttime) = 2019
GROUP BY rentdate
ORDER BY rentdate;
```

This Query groups all the rentals based on their start time (date) and gives data for number of rentals for the year 2019 along with their respective dates.

3. Group trip data by date and gender:

```
SELECT DATE(starttime) AS rentdate, gender, count(*) as volume
FROM trip
WHERE YEAR(starttime) = 2019
GROUP BY rentdate, gender
ORDER BY rentdate;
```

This query groups all the rental info based on the consumers' gender and the date of booking, again for the year 2019.

Section 6, Group 4

Python Data Source: We used the sportradar API to source our data for this project. We focused on Major League Baseball (MLB) games, particularly of Boston Red Sox in order to formulate our hypotheses of bluebikes which is also based in Boston.
Following is our Python code: -

```
import pandas as pd
import sportradar.MLB

key = "ABC"

client = sportradar.MLB.MLB(key)
date = [] # list of dates of Red Sox Home games
away_team=[] # list of away teams
status=[] # Status of game: whether postponed or not
audience=[] # Attendance during each game day
for year in range(2019,2020):
    sched = client.get_league_schedule(year, "REG") #Pulling MLB schedule for 2019
    sched = sched.json()
    for game in sched["games"]:
        if game["home"]["name"] == "Red Sox": # Pulling data for Red Sox games
            month = game["scheduled"].split("T")[0].split("-")[1] #extracting month of game day
            day = game["scheduled"].split("T")[0].split("-")[2] #extracting day of game day
            daysched = client.get_daily_schedule(year, int(month), int(day))
            daysched = daysched.json()
            for i in daysched["games"]:
                if i["home"]["name"] == "Red Sox":
                    date.append(game["scheduled"].split("T")[0])
                    away_team.append(game["away"]["name"])
                    try:
                        attendance = i["attendance"]
                        audience.append(attendance)
                        status.append(i["status"])
                    except KeyError:
                        attendance = 0
                        audience.append(attendance)
                        status.append(i["status"])

## Making a CSV file from data obtained

dict={'Schedule': date, "Away Team": away_team,'Attendance':audience, "Status": status }
df=pd.DataFrame(dict)
print(df)
df.to_csv('Schedule.csv')
```

The code uses the sportradar API to extract information in a csv format of the dates, away teams, attendance and status of all games played by Boston Red Sox for the year 2019. We used lists to store the information we needed as can be seen in the above code and finally concatenated the list as a dictionary before converting the same to a csv.

Finally, we aggregated all the csv files thus obtained into Tableau and joined tables based on startime (date of game/ rental) to analyse our hypotheses.

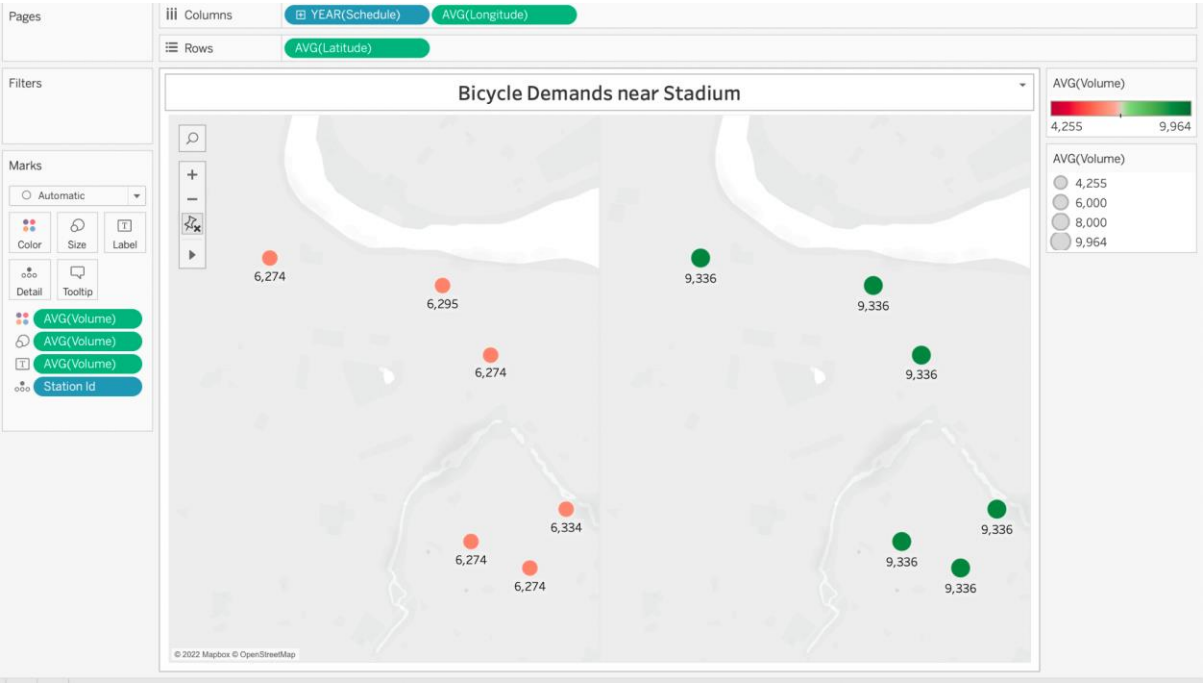
Section 6, Group 4

Hypothesis 2 Appendix Items:

SQL code:

```
SELECT DATE(starttime) AS rentdate, gender, count(*) as volume
FROM trip
WHERE YEAR(starttime) = 2019
GROUP BY rentdate, gender
ORDER BY rentdate;
```

Tableau Visual:



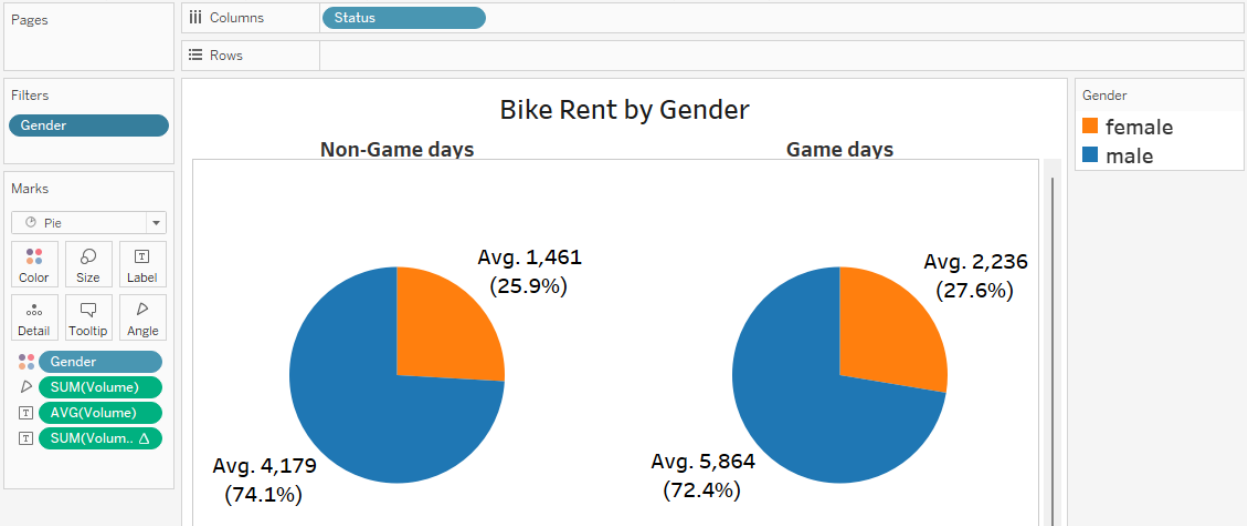
Hypothesis 3 Appendix Items:

SQL code:

```
SELECT DATE(starttime) AS rentdate, gender, count(*) as volume
FROM trip
WHERE YEAR(starttime) = 2019
GROUP BY rentdate, gender
ORDER BY rentdate;
```

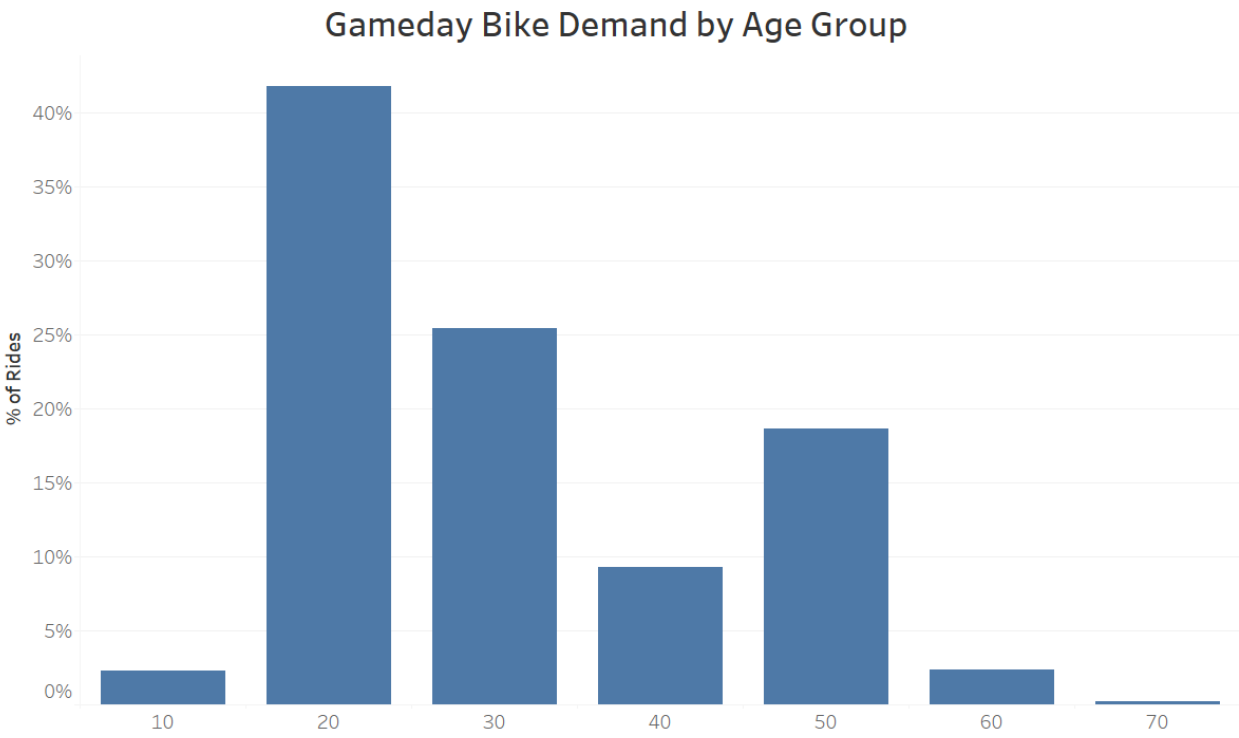
Tableau Visual:

Section 6, Group 4



Hypothesis 4 Appendix Items:

Tableau Visual:



Section 6, Group 4