

A Bike-sharing Business Analysis Project

First Edition

Qi Zhou
qiqizhou1996@gmail.com
18 March 2025

Introduction:

This project [1] demonstrates a fundamental data analysis process: asking questions, preparing datasets, cleaning data, analyzing and visualizing data, and sharing a compelling data story. The NY Citi bike-sharing program [2] aims to enhance its business strategy, thereby improving user experience.

Asking questions:

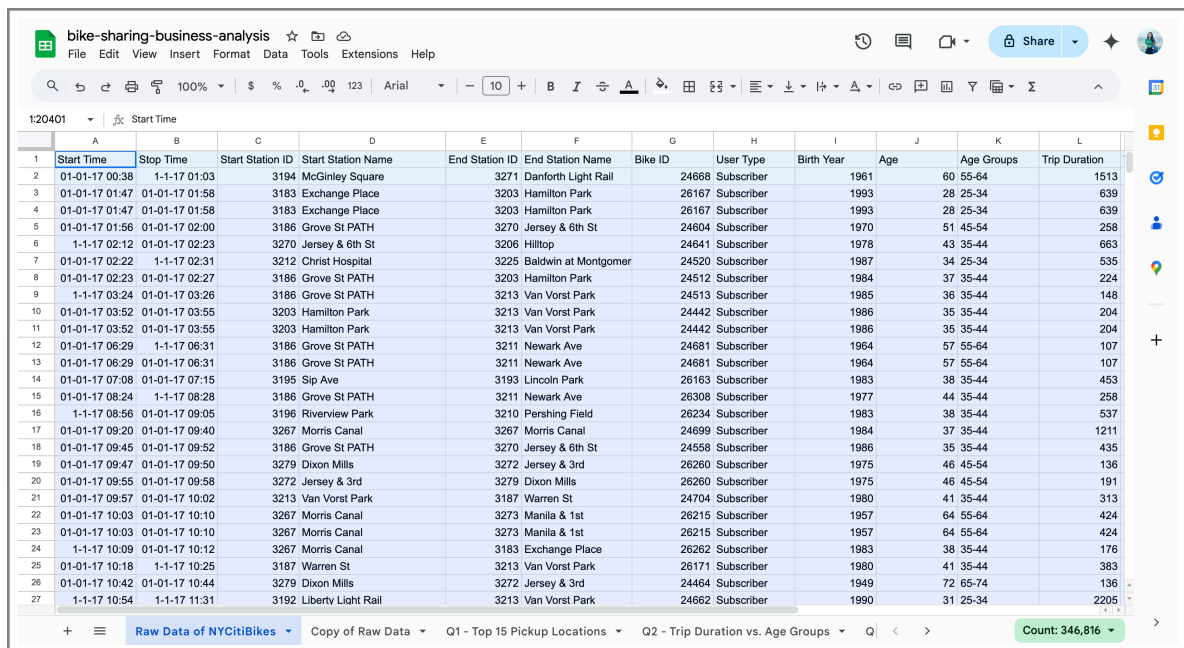
1. What are the most popular pick-up locations across the city for NY Citi Bike rental?
2. How does the average trip duration vary across different age groups?
3. Which age group rents the most bikes?
4. How does bike rental vary across the two user groups (one-time users vs. long-term subscribers) on different days of the week?
5. Does user age impact the average bike trip duration?

Preparing datasets:

The dataset [3] is chosen to answer the above questions, thus revealing how the bike-sharing service serves its users and to drive informed decision-making accordingly.

In this project, Google Sheets is the tool used to store, clean, and analyze the data. The raw data, which totals 20,400 records, spans three months, starting on January 1, 2017, and ending on March 31 of the same year, as shown in Figure 1.

Figure 1: Raw Data



	A	B	C	D	E	F	G	H	I	J	K	L
	Start Time	Stop Time	Start Station ID	Start Station Name	End Station ID	End Station Name	Bike ID	User Type	Birth Year	Age	Age Groups	Trip Duration
1	01-01-17 00:38	1-1-17 01:03	3194	McGinley Square	3271	Danforth Light Rail	24668	Subscriber	1961	60	55-64	1513
2	01-01-17 01:47	01-01-17 01:58	3183	Exchange Place	3203	Hamilton Park	26167	Subscriber	1993	28	25-34	639
3	01-01-17 01:47	01-01-17 01:58	3183	Exchange Place	3203	Hamilton Park	26167	Subscriber	1993	28	25-34	639
4	01-01-17 01:56	01-01-17 02:00	3186	Grove St PATH	3270	Jersey & 6th St	24604	Subscriber	1970	51	45-54	258
5	1-1-17 02:12	01-01-17 02:23	3270	Jersey & 6th St	3206	Hilltop	24641	Subscriber	1978	43	35-44	663
6	01-01-17 02:22	1-1-17 02:31	3212	Christ Hospital	3225	Baldwin at Montgomer	24520	Subscriber	1987	34	25-34	535
7	01-01-17 02:23	01-01-17 02:27	3186	Grove St PATH	3203	Hamilton Park	24512	Subscriber	1984	37	35-44	224
8	1-1-17 03:24	01-01-17 03:26	3186	Grove St PATH	3213	Van Vorst Park	24513	Subscriber	1985	36	35-44	148
9	01-01-17 03:52	01-01-17 03:55	3203	Hamilton Park	3213	Van Vorst Park	24442	Subscriber	1986	35	35-44	204
10	01-01-17 03:52	01-01-17 03:55	3203	Hamilton Park	3213	Van Vorst Park	24442	Subscriber	1986	35	35-44	204
11	01-01-17 06:29	1-1-17 06:31	3186	Grove St PATH	3211	Newark Ave	24681	Subscriber	1964	57	55-64	107
12	01-01-17 06:29	01-01-17 06:31	3186	Grove St PATH	3211	Newark Ave	24681	Subscriber	1964	57	55-64	107
13	01-01-17 07:08	01-01-17 07:15	3195	Sip Ave	3193	Lincoln Park	26163	Subscriber	1983	38	35-44	453
14	01-01-17 08:24	1-1-17 08:28	3186	Grove St PATH	3211	Newark Ave	26308	Subscriber	1977	44	35-44	258
15	1-1-17 08:56	01-01-17 09:05	3196	Riverview Park	3210	Pershing Field	26234	Subscriber	1983	38	35-44	537
16	01-01-17 09:20	01-01-17 09:40	3267	Morris Canal	3267	Morris Canal	24699	Subscriber	1984	37	35-44	1211
17	01-01-17 09:45	01-01-17 09:52	3186	Grove St PATH	3270	Jersey & 6th St	24558	Subscriber	1986	35	35-44	435
18	01-01-17 09:47	01-01-17 09:50	3279	Dixon Mills	3272	Jersey & 3rd	26260	Subscriber	1975	46	45-54	136
19	01-01-17 09:55	01-01-17 09:58	3272	Jersey & 3rd	3279	Dixon Mills	26260	Subscriber	1975	46	45-54	191
20	01-01-17 09:57	01-01-17 10:02	3213	Van Vorst Park	3187	Warren St	24704	Subscriber	1980	41	35-44	313
21	01-01-17 10:03	01-01-17 10:10	3267	Morris Canal	3273	Manila & 1st	26215	Subscriber	1957	64	55-64	424
22	01-01-17 10:03	01-01-17 10:10	3267	Morris Canal	3273	Manila & 1st	26215	Subscriber	1957	64	55-64	424
23	1-1-17 10:09	01-01-17 10:12	3267	Morris Canal	3183	Exchange Place	26262	Subscriber	1983	38	35-44	176
24	01-01-17 10:18	1-1-17 10:25	3187	Warren St	3213	Van Vorst Park	26171	Subscriber	1980	41	35-44	383
25	01-01-17 10:42	01-01-17 10:44	3279	Dixon Mills	3272	Jersey & 3rd	24464	Subscriber	1949	72	65-74	136
26	1-1-17 10:54	1-1-17 11:31	3192	Liberty Light Rail	3213	Van Vorst Park	24662	Subscriber	1990	31	25-34	2205

Cleaning data:

After obtaining the raw data, it is first backed up. Then, the backup data is cleaned, and the raw data remains intact to ensure traceability and reproducibility of subsequent cleanup operations. Regarding this data, the cleanup includes three steps - identifying and removing duplicates, trimming whitespaces, and handling missing data points as well as outliers.

Compared with the raw data, the cleaned data totals 16,843 records, as shown in Figure 2. Besides, the attributes are highlighted with a light green background and bold font so as to help understand the table quickly and easily.

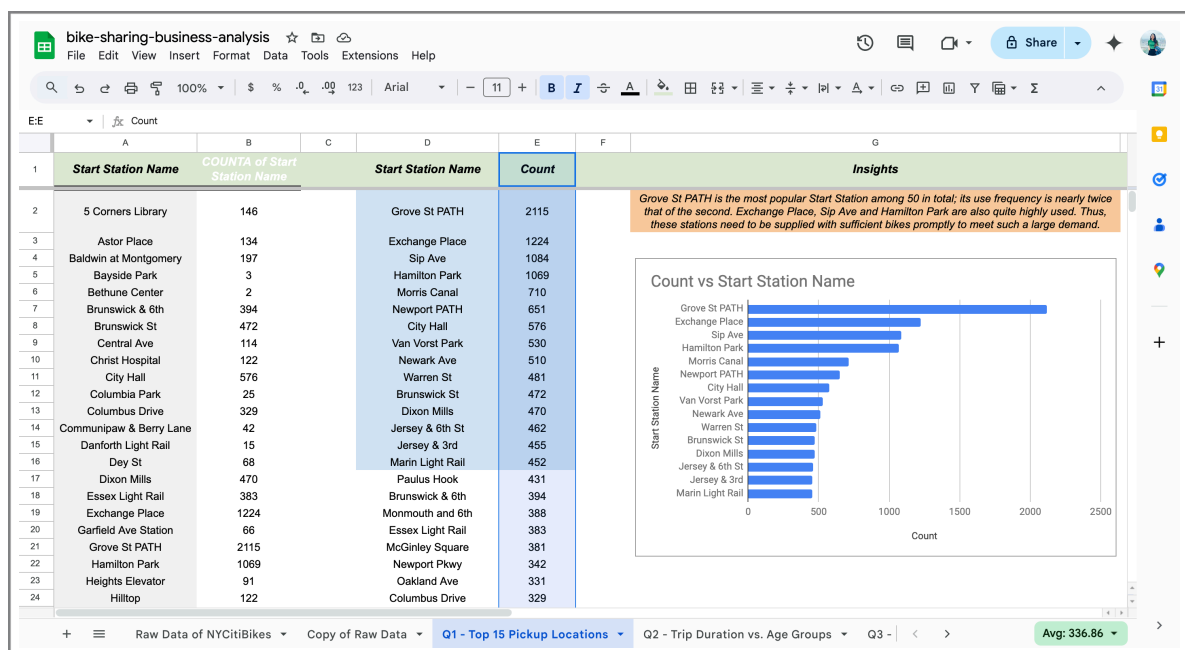
Figure 2: Cleanup on Backup Data

	A	B	C	D	E	F	G	H	I	J	K	L
1	Start Time	Stop Time	Start Station ID	Start Station Name	End Station ID	End Station Name	Bike ID	User Type	Birth Year	Age	Age Groups	Trip Duration
2	10-3-17 17:39	13-03-17 08:12	3185	City Hall	3277	Communipaw & Berry La	24462	Subscriber	1995	26	25-34	221604
3	16-01-17 16:36	18-1-17 08:58	3203	Hamilton Park	3213	Van Vorst Park	24716	Subscriber	1983	38	35-44	145342
4	17-02-17 22:06	19-2-17 14:27	3267	Morris Canal	3183	Exchange Place	24418	Subscriber	1941	80	75+	145296
5	15-01-17 17:20	16-01-17 18:27	3202	Newport PATH	3185	City Hall	24495	Subscriber	1970	51	45-54	90394
6	27-03-17 18:30	28-3-17 19:22	3186	Grove St PATH	3192	Liberty Light Rail	26170	Subscriber	1957	64	55-64	89479
7	15-1-17 11:39	16-01-17 11:26	3203	Hamilton Park	3205	JC Medical Center	26254	Subscriber	1989	32	25-34	85624
8	25-02-17 13:14	26-2-17 12:44	3192	Liberty Light Rail	3275	Columbus Drive	26250	One-time user	1984	37	35-44	84618
9	17-2-17 17:30	18-02-17 15:41	3269	Brunswick & 6th	3273	Manila & 1st	26208	Subscriber	1984	37	35-44	79860
10	25-03-17 14:18	26-03-17 11:15	3203	Hamilton Park	3206	Hilltop	24406	One-time user	1984	37	35-44	75410
11	25-02-17 16:53	26-2-17 12:52	3185	City Hall	3278	Monmouth and 6th	24395	Subscriber	1981	40	35-44	71910
12	29-03-17 13:36	30-03-17 06:58	3183	Exchange Place	3211	Newark Ave	29266	One-time user	1984	37	35-44	62507
13	24-02-17 19:42	25-02-17 10:17	3186	Grove St PATH	3280	Astor Place	24568	Subscriber	1985	36	35-44	52513
14	25-02-17 15:32	26-2-17 04:22	3209	Brunswick St	3215	Central Ave	24607	Subscriber	1994	27	25-34	46196
15	21-3-17 20:10	22-03-17 08:21	3267	Morris Canal	3186	Grove St PATH	24544	One-time user	1984	37	35-44	43871
16	5-1-17 23:16	06-01-17 08:33	3270	Jersey & 6th St	3186	Grove St PATH	26230	Subscriber	1988	33	25-34	33426
17	22-3-17 09:52	22-3-17 18:31	3267	Morris Canal	3183	Exchange Place	24724	Subscriber	1983	38	35-44	31177
18	13-3-17 12:20	13-03-17 20:50	3207	Oakland Ave	3183	Exchange Place	24695	Subscriber	1976	45	45-54	30620
19	09-03-17 00:35	09-03-17 08:30	3185	City Hall	3185	City Hall	26315	Subscriber	1983	38	35-44	28480
20	08-03-17 08:23	8-3-17 14:17	3275	Columbus Drive	3275	Columbus Drive	24571	One-time user	1984	37	35-44	21217
21	22-2-17 17:42	22-02-17 23:04	3183	Exchange Place	3275	Columbus Drive	24459	Subscriber	1978	43	35-44	19298
22	14-01-17 11:21	14-01-17 16:33	3199	Newport Pkwy	3199	Newport Pkwy	26238	One-time user	1984	37	35-44	18710
23	14-2-17 17:31	14-02-17 22:41	3207	Oakland Ave	3211	Newark Ave	24687	Subscriber	1970	51	45-54	18610
24	25-02-17 15:03	25-2-17 20:04	3186	Grove St PATH	3273	Manila & 1st	26197	One-time user	1984	37	35-44	18055
25	24-02-17 18:28	24-2-17 22:49	3202	Newport PATH	3202	Newport PATH	26164	One-time user	1984	37	35-44	15676
26	16-01-17 11:19	16-01-17 15:31	3202	Newport PATH	3187	Warren St	24476	Subscriber	1982	39	35-44	15167

Analyzing & Visualizing data:

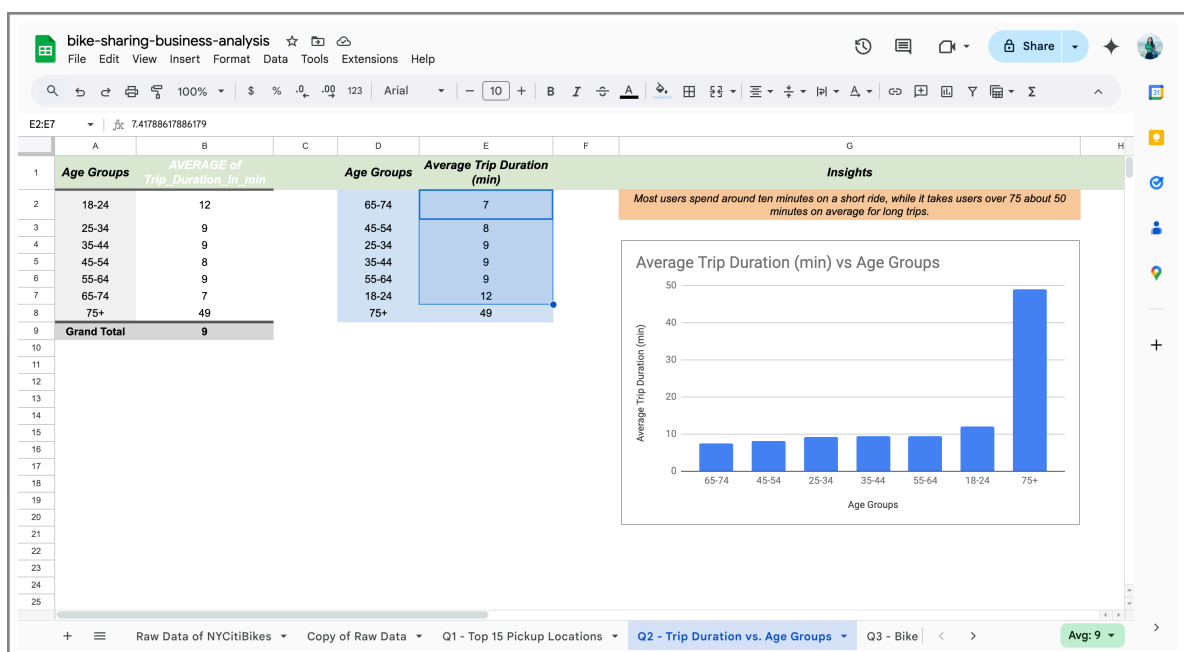
To answer the first question, a pivot table is created to summarize the use frequency of each starting station. After sorting in descending order, Grove St PATH station is the most popular starting station among the 50 stations, with 2,115 out of 16,843 data records (note that the average is 337). Moreover, as the bar chart in Figure 3 suggests, Exchange Place, Sip Ave, and Hamilton Park stations are quite highly used. Thus, a possible business strategy is supplying these stations with sufficient bikes promptly to meet the large demand.

Figure 3: Q1 - Top 15 Pickup Locations



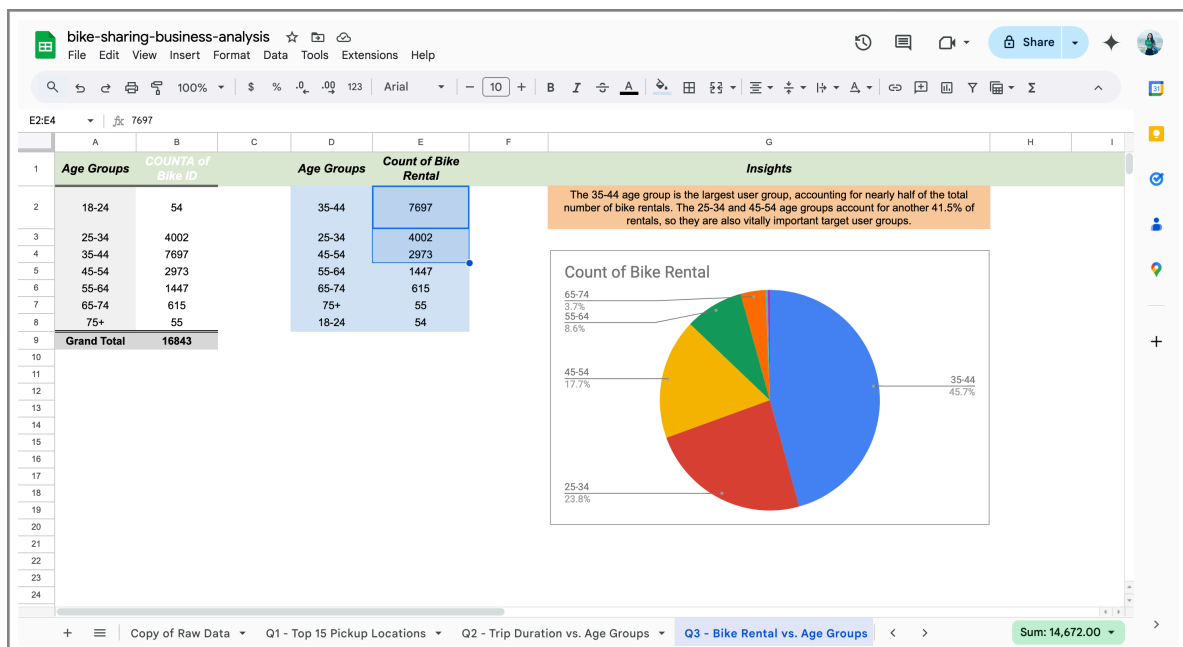
For the second question, a pivot table summarizes the average trip duration for seven age groups (18-24, 25-34, 35-44, 45-54, 55-64, 65-74 and 75+). As the column chart in Figure 4 suggests, most users aged 18 to 74 spend an average of 9 minutes on a short ride, while users over 75 spend an average of 50 minutes on long trips. Thus, the business strategy would focus on the ten-minute ride.

Figure 4: Q2 - Trip Duration vs. Age Groups



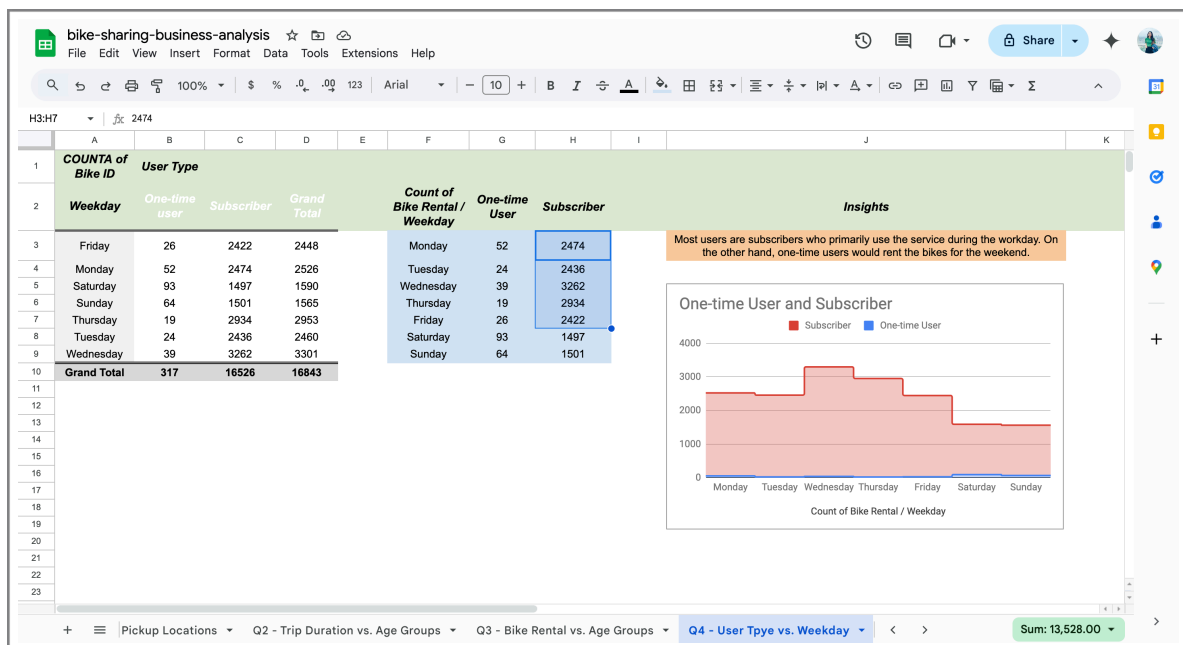
For the third question, a pivot table summarizes the number of bike rentals for the seven age groups. As the pie chart in Figure 5 suggests, the 35-44 age group is the largest user group, with 7,697 of 16,843 rentals, accounting for nearly half of the total. In addition, the 25-34 and 45-54 age groups account for another 41.5% of rentals, so they are also vitally important user groups. Thus, the business strategy would mainly target users aged 25 to 54.

Figure 5: Q3 - Bike Rental vs. Age Groups



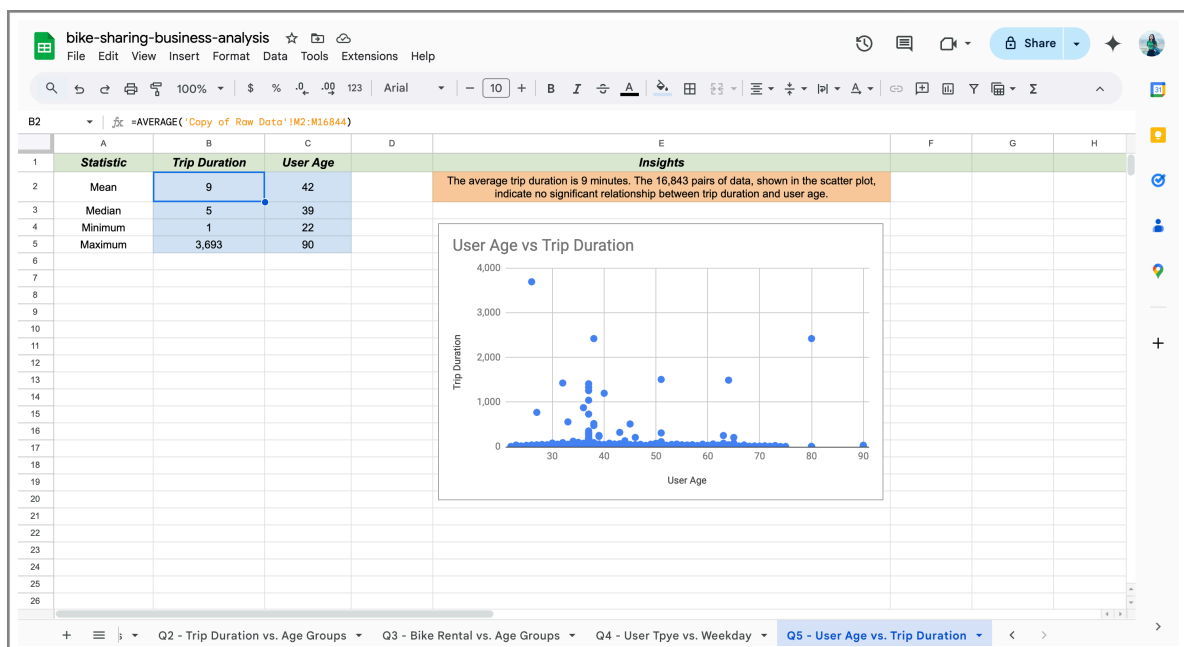
For the fourth question, a pivot table summarizes the number of bike rentals per weekday by one-time users and subscribers. As the stacked stepped area chart in Figure 6 suggests, most users are subscribers who primarily use the service during the workday, with 13,528 of 16,843 rentals. On the other hand, one-time users tend to rent bikes for the weekend. Thus, the business strategy would greatly retain subscribers and try to attract one-time users on the weekends.

Figure 6: Q4 - User Type vs. Weekday



Regarding the last question, descriptive statistics, in which the average trip duration is 9 minutes (shown in Figure 7), are used to generate a general intuitive understanding of the whole 16,843 data records and help effectively eliminate outliers. As the scatter plot in Figure 7 suggests, the 16,843 pairs of data indicate no significant relationship between trip duration and user age. Thus, it seems that user age does not impact the average bike trip duration. For further questions, more data is needed to conduct a detailed analysis.

Figure 7: Q5 - User Age vs. Trip Duration



Data storytelling:

This project leverages descriptive analysis to understand how the NY Citi bike-sharing program operates, plus its customer behaviour, thus providing specific solutions for upgrading the business strategy.

The raw dataset contains 20,400 records, spanning three months from January 1 to March 31, 2017. After data cleaning, the dataset retains 16,843 records. According to the cleaned data, there are 50 starting stations, and the average number of services provided by the stations over these three months is 337. However, four stations - Grove St PATH, Exchange Place, Sip Ave, and Hamilton Park - each provide over 1,000 times of service. Therefore, these stations should be supplied with sufficient bikes promptly to meet this high demand. Additionally, more than 80% of customers are subscribers, primarily aged between 25 and 54, who are accustomed to renting bikes for ten-minute short rides during workdays. Thus, the business strategy should focus on retaining these customers through subscriber benefits and improved riding experiences.

Acknowledgements:

[1] Thanks to the CareerFoundry Data Analytics for Beginners Course. For details, please see careerfoundry.com

[2] For details, please see citibikenyc.com

[3] The original datasets are from kaggle.com

[4] I am very grateful to the Google Data Analytics Professional Certificate for helping me systematically and accurately develop a knowledge framework, gain practical experience, and cultivate an analytical mindset in data analytics. This is incredibly useful for me independently achieving this business analysis project.