Chicago’s Shared Bikes: How Big Data Technology Can Assess Ridership

Carrie D. Keenan

Abstract—Big Data analytics is a growing field. Not only because there are more actual users utilizing Big Data and Big Data Ecosystems, but because the amount of data, the types of data, and speed at which these data are changing is also growing. This project utilizes the functionality of Cloudera’s Hadoop Distributed File System (HDFS), Hive, and MapReduce capabilities to examine a data set of one year’s worth of information about bike share riders in Chicago and the way weather affects ridership to answer the following questions:

Do annual subscribers ride their bikes in the rain more than daily riders?

Are there any trends in ridership type and season or temperature?

Is the mean trip duration greater on days where the weather is ‘nicer’ with little to no rain and moderate temperatures?

Cloudera and the Hadoop Ecosystem provide a platform to query and analyze the data to find these answers while also utilizing non-Hadoop options to temper the learning curve.

I. INTRODUCTION

Divvy is the bicycle sharing program in the City of Chicago. Launched in June 2013, the program currently has over 4500 bikes and more than 475 bike stations across the city. It was developed as part of a federal grant through Chicago’s Department of Transportation to “promote economic recovery, reduce traffic congestion and improve air quality” under Mayor Rahm Emanuel [6]. Divvy is operated by Motivate which, in addition to Chicago, also operates the bike share programs in Washington, DC, Boston, New York, San Francisco, Columbus, OH, Chattanooga, TN, and Melbourne, Australia.

The bike-share system allows users to utilize a bike without having to own or maintain one. The bicycles are docked at one of the bike stations, used temporarily, and then returned to another (or the same) bike station. The ‘Chicago Blue’ color bikes are unisex, semi-adjustable, and heavy duty. In addition to leisurely rides in the city, the bikes can also be used as a means of commuting. Divvy offers both daily rental passes and annual memberships. Annual memberships run $75 with additional options for students and those with low-incomes ($55/$5). Daily use costs $9.50. Regardless of renter type, the bikes must be re-docked every 30 minutes; you can rent all day, but you need to exchange your bike at a new station after a half-hour of use.

The integration of Divvy into Chicago is not a standalone project. In the past 5 years, major improvements have, and are, being made to increase access to both active transit options as well as better public transit.

Currently, there are just over 200 total miles of bike lanes in the whole of Chicago. In 2006 a plan was developed by Mayor Rahm Emanuel, the Bike 2015 Plan, which proposed two primary objectives. The first is “to increase bicycle use, so that 5 percent of all trips less than five miles are by bicycle” [4]. The second aim is to reduce the number of bicycle injuries by 50 percent from current levels. Although the plan did not reach the goal of 500-miles of bikeways by 2015, it did greatly improve existing lanes and increase the overall bike path and bike lane mileage from just over 125 miles to over 200, a 60% increase. This addition of better routes has also led to an increase in bikers from 0.5% in 2000 to 1.3% in 2010 and has continued to rise making Chicago the 5th most bicycle-friendly city in the United States.

With the investment in the infrastructure of bike safety elements like bike lanes and in the bike share program, there is an opportunity to examine how biking in the city is affected by environmental factors. In particular, is weather a factor in whether or not people ride? Are people less likely to ride in the rain? Divvy has two types of riders: those that buy an annual pass and those that use a day pass. Is there a difference in the type of rider and how these environmental factors influence the amount of ridership?

Based on my personal experience I am less likely to ride in the cold than I am during nice weather. But what is ‘nice’? Can weather that is too hot have as much of an influence as weather that is too cold? Beyond that, is precipitation a factor? While I would argue, based on my own bias, that Divvy bikers are not the same as ‘real’ bikers, their use of bikes annually does provide an incredibly rich source to examine the effect on biking patterns in Chicago.

Using the Hadoop system, Cloudera, I will examine all divvy rides from July 2014-June 2015 and the effects of the weather on those rides and rider types. Some questions I will explore are:

- Do annual subscribers ride their bikes in the rain more than daily riders?
- Are there any trends in ridership type and season or temperature?
- Is the mean trip duration greater on days where the weather is ‘nicer’ with little to no rain and moderate temperatures?
Overall, I hope to establish some trends with this large dataset and examine the connecting factors that account for these trends.

II. USING BIG DATA ECOSYSTEMS

A. Big Data Ecosystem

The Big Data ecosystem can be broken down into four main categories: data, storage and computing infrastructure, transformations and analytics, and data visualization and delivery. The first category, data, includes both the sources and collection which is to say all the various types of data and how they are collected. Data can come from traditional enterprise data, public data, machine sensor data, or social data like Twitter and Facebook. For this project, I used public data from two main source, Divvy and the National Oceanic and Atmospheric Administration (NOAA) as well as some self-created and modified data.

The second category is data storage and computing infrastructure. Once the data is acquired, it needs to be stored and made accessible. This is where Hadoop comes in the form of data storage clusters that can handle the large volumes of data, but also includes database software, NoSQL platforms and Massively Parallel Processing databases that have the capacity to process the data. For the purposes of this project I did not gain multi-node clusters that might be required for larger or more complex storage and processing. But, in comparison with other systems I have used this data set with in the past, like Weka, Hadoop can handle over a million records without freezing or crashing.

The third category is data transformation and analytics. The third category is where the data is really put to use as it is processed via MapReduce, Pig and Hive to analyze for faster analysis of these large data sets. Once processed the data can be analyzed using different methods including applying machine learning algorithms and data mining techniques. In this project, I used Hive solely as I felt the most comfortable with its syntax. Overall the querying of data, while slow, was fairly straight forward and allowed me to obtain answers to my questions.

Finally, the fourth category is the data visualization and delivery. This component takes the analyzed data and makes it presentable, like Tableau. There are many different products that have been created to make each of these areas of big data run faster, smarter, and more efficiently. For this project I was not able to explore the options offered by the Hadoop ecosystem. While I did experiment with Hive’s native display tool, I found it cumbersome and eventually I reverted back to a more familiar, if basic, data display tool, Excel.

One of the biggest advantages of Hadoop Distributed File Structure (HDFS) and Hadoop in comparison to traditional information technology systems is that is can take multiple forms of data – Relational Database Management Systems (RDBMS) and nonRDBMS – and integrate it together for analysis. Hadoop can handle large volumes of data and process it quickly. The ability to take the data into Hadoop, manipulate it, query it, and then return it to a traditional RDBMS like Excel makes it very useful especially as technologies are being developed and there is a learning curve. The two can work together to evaluate the information, at this point it is not an and/or problem. I found it incredibly useful to be able to jump between Stata, Hive, and Excel to find the answers to my questions within the data set.

B. Hadoop Ecosystem

I created this data flow to show the Hadoop Ecosystem, Fig. 1. For this project, I utilized four pieces of this system: data storage, data processing, data programming, and data presentation.

Within the Hadoop Ecosystem of Cloudera, these pieces work together to compile, store, manage, and analyze large data sets. Although all of these components can work together, a user does not need, or necessarily want, to use all of them. Hadoop and Cloudera do not make you choose. But as you can see, the whole ecosystem works together to assist in processing big data.

- Hadoop Distributed File System (HDFS): An open source software stack that runs on a cluster of machines and provides scalability, redundancy, fault-tolerance, cross-platform compatibility and storage-compute in one environment.
- HBASE: A non-RDBMS that runs on top of HDFS. It scales linearly so it can handle very large data sets with different structures and schemas.
- MapReduce: For processing distributed computations on the data clusters.
- YARN (Yet Another Resource Negotiator) is MapReduce 2.0
- Zookeeper: Coordinates services and tasks between nodes.
- Pig: High level platform and programming language for analyzing and handling the large data sets.
- Hive: Provides data summarization, analytics, and queries in a SQL-like language, HiveQL.

For my project, I used Hive because I am more familiar with the syntax and felt like I was able to obtain the answers to the questions I was investigating. I also benefitted from Cloudera’s ability to take query results, copy them to a .csv and transfer that to Excel. While Cloudera has some functions for data presentation, and Hadoop has several focused utilities, like Tableau, I found that Excel’s capabilities met my needs.

In addition, while working with my data set, I struggled with some of the fields, namely the date, in being able to manipulate the data within Cloudera. I found that combining my knowledge of Stata to pre-process the data helped immensely in being able to transfer the data to Cloudera and begin analysis. Given more time and experience, I believe that all of my processing could have been accomplished within Cloudera, but for a start, I was satisfied with the results and my experience within Cloudera’s Hadoop ecosystem.
III. MATERIALS AND METHODS

A. Data Sets

a) Divvy data

As a publicly funded entity, Divvy makes de-identified datasets available to the public. Data is release bi-annually at the end of quarter 4 and the end of quarter 2 with data going back to quarter 3 of 2013, when Divvy began operation in Chicago. There are a total of four datasets available, with the last set in quarter 2 of 2015.

Each data set contains a trip id, trip start and end times, start and end stations, rider type, and member information. There are over 475 solar-powered stations across the city and geo-coordinate location information is available for each station. There are three rider types: customer, subscriber, and dependent. Customers are single riders, those without an annual membership. Subscribers and dependents are annual membership holders and there is additional information for them including birth year and gender.

For this research, I used the data from quarter 3 and 4 (Q3-Q4) 2014 and Q1-Q2 2015, July 1, 2014-June 30, 2015. This represents one fiscal year. There were a total of 2,645,174 instances in the combined initial data sets.

b) National Oceanic and Atmospheric Administration (NOAA) data

NOAA offers several datasets of weather-related data including wind speed and direction, humidity, precipitation, and daily temperatures. For this research, I utilized the daily temperature and precipitation data. Of the data collection sites NOAA has in Chicago, the most complete dataset for my needs was collected from Chicago’s Midway Airport (MDW). The data include maximum temperature (TMAX), minimum temperature (TMIN), and precipitation (PRCP). The precipitation is in tenths of millimeters and the temperatures are in Celsius degrees to tenths.

I obtained the data covering the same dates as the divvy data. There were 365 data points.

c) Self-created data

I created a variable Season. The season variable divided the data into quarters for Fall (1), Winter (2), Spring (3), and Summer (4). Fall is defined as September 22, 2014-December 21, 2014, Winter as December 22, 2014-March 21, 2015, Spring as March 22, 2015-June 21, 2015, and Summer as July 1, 2014-September 21, 2014 and June 22, 2015-June 30, 2015.

In addition to the season information, I created a few new fields by manipulating the existing data. First, I converted the weather into an average in Fahrenheit using the TMAX and TMIN fields. Second, I transformed the precipitation into inches. I also created fields to indicate rain/no rain, annual member/daily member, and age.

B. Data Processing

a) Metastore Tables and File Browser

Using File Browser and Metastore Tables, the files needed for my analysis were uploaded. Once uploaded, they were converted to tables. I created five Divvy tables for Quarters 3 and 4 of 2014 and Quarters 1 and 2 from 2015. Q3 2014 was in two separate files. Additional other files were also processed into tables via Metastore and file browser including divvy_noaa.

b) Hive

In addition to using Hive for analysis, Hive was also used for processing. I used Hive to combine all Divvy tables into a single table, divvyall, with 2,645,174 records. After trial and error within Hive, I used Stata to convert the date field. I reimported the divvy and weather data and still had difficulties processing the data due to the date field. Due to time constraints, I decided to process the remaining data in Stata so I could spend more time working with the data set. The complete dataset was brought into Cloudera as ‘divvy_complete’, which still has the expected 2,645,174 records. To utilize the birth year field as ‘age’, I went back to stata and created an age field, which I ten joined to the ‘divvy_complete’ file in Hive.

c) Stata

Ideally, I would have preferred to do all data processing in Cloudera. Unfortunately, I had difficulties with manipulating the data, in particular the date field. Using the Stat/Transfer tool, I converted the Divvy files into .dta Stata files. Once they were individual .dta files I imported the first into Stata and then appended each one into a single file: divvyall.dta. This is the same file created with Hive. From there I manipulated the date field and exported the data into a .csv ‘divvyall_date’. When this was not effective in allowing me to merge data within Hive, I brought it back in to Stata and made a few more necessary modifications. I created a binary field, ‘prcp_yn’, for days that have precipitation or not and ‘usertype’ for whether the riders are annual members or daily customers. I also created the following fields:
- ‘av_temp’ which shows the mean of the maximum and minimum temperature
- ‘avtemp_f’ which is the temperature in Fahrenheit, converted from tenths of degrees Celsius
- ‘prcp_in’ which is converts the precipitation from tenths of centimeters to inches
- ‘trip_min’ which calculates the trip duration into minutes
- ‘age’ was created later, subtracting the birth year from the current year, 2015.

In Stata, the self-created the season field was constructed based on the date. Once I was satisfied with the cleanliness of the data using Stata, the project moved into the experimentation phase within Cloudera.

d) Excel

Excel was used for creating more visibly appropriate graphs from the data produced from Cloudera. Overall, the graphs are good for the type of analysis. In future, it would be interesting
to do this step with Tableau or another program that is primarily for data presentation.

IV. EXPERIMENTS AND RESULTS

A. Use Test 1 - Do annual subscribers ride their bikes in the rain more than daily riders?

For my first test, I wanted to see if there was a difference in ridership when there is rain. My hypothesis is that those who purchased annual passes are more likely to be committed to biking and will therefore be more likely to ride in the rain.

I ran the following query: select count(*) from divvy_complete where user=1 and prcp_yn=1, group by user. This captured a count of annual pass and daily riders on days where there was rain. This provided a count of 589,311 for riders with an annual pass (user=1) and 243,479 for daily riders (user=0). There are 345,832 more annual riders that took rides during precipitation, over double. In comparison, I also examined the number of days that had rain. Of the 365 days accounted for in the data set, there were 117 days with precipitation, or 32.1%. Of the 2,645,174 total rides, the number of rides on days with precipitation is 832,790 or 31.5%. Whether it has any impact on Divvy’s business strategies, I find it interesting that roughly 1/3 of all rides take place on the 1/3 days where there is rain. Regardless, it is clear that more of these precipitation rides are by annual riders, 22.3% compared to 9.2% of daily riders.

Seeing the pure count is just one way to evaluate the data. Hive provides additional tools to examine the weather. To illustrate the count of annual riders by the amount of precipitation, I pulled the count by user and grouped by precipitation in inches. I chose to count by date because, I wanted to see the number of riders on a day with precipitation and compare it with the amount of precipitation. This provides a graph with the amount of precipitation on the x-axis and the count of riders with an annual pass on the y-axis. Unfortunately, although Hive provides a simple graphical representation, I found it hard to see due to the number of data points and the fact that you cannot modify the axis text. In order to do this, I decided to use an outside source for more detailed graph manipulation.

I copied this data into .csv and imported it into Excel. I repeated the process with the non-annual riders and added those numbers to the excel file. These were then graphed to compare the number of rider types by amount of rain. I removed the users when there was no precipitation to make the graph more legible. There were 1,252,775 annual riders on days with no precipitation and 559,609 daily riders. From the remaining data, all days with precipitation, there is a trend of more riders when there is less rain, Fig. 2. There is some ridership regardless of amount of precipitation, but after about .5 inches of precipitation there are less riders in general.

B. Test 2 - Are there any trends in ridership type and season or temperature?

For test number 2, I pulled a similar query as test 1, replacing ‘avtemp_f’ for ‘prcp_in’ to determine the number of daily/annual riders by average daily temperature. Grouping the temperature allowed me to look at a certain single temperature across months and seasons with rider type. I did this because you have multiple days with the same temperature.

In this initial evaluation, you can see a defined trend in the more riders in ‘warmer’ temperatures and less in ‘colder’ temperatures, Fig. 3. There is a significant spike of annual riders at 70.52 degrees with 85,095 riders; daily riders also have a higher count at this temperature with 32,635. The median number of riders for each temperature grouping is 6,603 and 1,295 for annual and daily riders respectively. The average for annual and daily riders is 10,011 and 4,365, respectively. The number of riders for both annual and daily riders that ride on days when the temperature is 70.52, are both well above both the median and mean for their rider types.

To calculate average temperature, I used the maximum and minimum temperatures obtained from the NOAA data set, averaging the two and converting them to Fahrenheit. Although this does restrict some of the complexity of weather, for this analysis it made sense to simplify it to one daily average temperature. Due to the high number of riders, eight times the average for annual riders, on days when the temperature is 70.52, I used Hive to calculate a count of days that met that temperature. You can see that there are 9 dates that fall into this temperature average, 5 are in summer, 2 in spring, and 2 in fall, Fig. 4. The two in fall have the lowest count of riders.

This trend in seasonality led me to see how the information above relates to the season. I ran the same query, but I included season in the grouping. Occasionally, there is a very warm day in winter and a cold day in summer. I wanted to examine if seasonality affected the number of riders more or less than the actual temperature. In other words, are riders likely to ride in the fall/winter when there happens to be a warmer day, or have they put away their biking shoes for the season. Again, for clarity, I brought the data into Excel to modify the graph.

I graphed it two ways. First, by season. I found this a little misleading because you can’t see the trend in temperature day to day across the season, Fig. 5. You can only see the minimum to maximum average temperature within the season. I re-graphed the data, adding the additional grouping of date. When I graphed it in Excel, you can now see the trend of temperature day to day across the entire year, Fig. 6.

When examining the second of these two graphs you can see clear indications of the weather affecting the amount of ridership. I have indicated with pink arrows on the graph on page 20 two areas, October 4, 2014 and May 29, 2015. You will notice that at these two points, there is a clear drop in both ridership and the temperature. In general you can see this trend, when the temperature drops, ridership drops and when the temperature rises, so does the quantity of riders. While these are not 100% correlated, it is interesting to look at the comparison of these two.

I was interested in the change over time by season. It looks as though, although ridership overall does drop significantly in
the winter, especially the daily rider rates, the temperature affected count follows the overall trend. When comparing the annual and daily riders, the graph shows that the annual riders follow the trend more closely, whereas the daily riders seem to react slightly differently. Indicated in black on page 20 you can see the daily riders reacting in opposition to both the annual riders and the temperature direction.

C. Test 3- Is the mean trip duration greater on days where the weather is ‘nicer’ with little to no rain and moderate temperatures?

To run the analysis for test 3, I needed to define ‘nice’ weather. I found a survey “Best temp for 20 mile ride” that established an appropriate range for good biking weather (Bike Forums 2007). The sample was small (n=126), but it made a clear divide. There were 35.7% voted for 60-69 degrees Fahrenheit and 33.3% voted for 70-79 degrees. For this test, I opted to combine these two and look at the range of 60-79 degrees as the best range for riding, considering over half of the riders surveyed fell into this range.

In addition to temperature, it is likely more ideal to ride when there is no rain. Examining the data from test 1, you can see that there is still a significant number of riders at very low levels of precipitation. On days when there are between 0.01 and 0.10 inches of precipitation there are 247, 654 annual riders and 98,144 daily riders. To include those rides in this analysis, I am going to define ‘nice’ weather as 0-0.10 inches of precipitation and an average temperature of 60-79.

First I looked at the average trip duration within this specific range of nice weather. The average is 16.99 minutes. In comparison, the overall average length of a ride on days without rain is 17.01, compared with rainy days of 16.57. The length of the nice weather is only two-hundredths of a minute different from the average on days with no rain.

To examine this graphically, I ran the data pulling the overall averages by temperature for days when there was less than 0.1 inches of precipitation, grouped by the average temperature. Again, I moved this into excel to graph it, Fig. 7. Overall there is a trend moving from shorter rides during colder temperatures to longer rides at warmer temperatures, however, there are a few outliers. There is a significant spike at 10.22, 15.62, 19.22, 20.57, and 21.20. Why? I ran an analysis to look at these lower temperatures, the riders and the precipitation.

The majority of days with a temperature less than 21.3 degrees add no precipitation. Recall that precipitation levels from 0-0.1 inches were included as ‘nice’ weather. In this set, there is only one day with more than 0.1 inches. Overall, though, there doesn’t seem to be a correlation between precipitation and the spikes in longer rides, Fig. 8. What stands out the most in the data is that the spikes are primarily from daily Divvy riders rather than annual riders.

In previous tests, it is clear that there are many more annual riders across the year than daily riders, but in this instance the daily riders seem to be affecting the average trip duration more than the annual riders. Were there several riders that rode a long duration, or was this a result of a few. I removed the precipitation and added a count by ridership. As you can see, it appears that these spikes are the result of only a few daily riders, Fig. 9. The average duration for daily riders on the days with the two dramatic spikes are 116.4 minutes with a temperature of 2.21 and 96.41 at temperatures of 15.62. At the same temperatures for annual riders, the average duration is 8.7 and 33.4. The 15.62 temperature is one of the clear spikes on the ‘nice temperature’ graph. Part of the difference in profiles between the two graphs is due to averaging by user type as opposed to by overall. In all, the data shows that on these cold days, only a few riders led to these sharp rises in trip duration for daily riders, 14 and 13 riders for the two spike days. Based on the need to renew a bike every 30 minutes, I would assume that this spike is due to a few of these riders forgetting to return the bikes within the 30-minute window and continuing to ride.

V. CONCLUSION

As a result of this analysis, my conclusion is that in general, subscribers are more likely to ride their bikes in the winter as opposed to the daily user. They ride more when it is cold and when there is precipitation. Generally, however, all riders ride in the spring and summer more than the fall and winter. Riders ride more on days when it is not precipitating.

One interesting side note I observed in the data is that the overall average ride is 16.87 minutes. Divvy requires bikes to be returned to a station every 30 minutes. With the average just over half of the time allowed, it makes me wonder how often people are worried about returning the bike late and instead opt for more frequent exchanges. From a business perspective, Divvy might consider increasing this time to attract customers who shy away from the exchange part of the Divvy model.

In test 3, I examined the drip duration and weather and saw an overall correlation. On days that had a higher temperature, in general, the rides were longer and vice versa. Regardless of the temperature, however, the rides averaged within the 30-minute rental time. The exception to this was the collection of days in which the temperature was 15.62 degrees F. Here you see a small amount of daily bike riders (n=13) skewing the data for days of this temperature by averaging over an hour and a half. This spike is an outlier.

Test 1 showed that the precipitation plays a role in ridership, but not nearly as much as I would have expected. While on days with less precipitation there are clearly more riders overall, you do see nearly the same number of riders on a day with 2-3 inches as on days when there is only 1/3 inch of precipitation. The one area that is not fully exposed by this analysis is time. If there was a heavy thunderstorm that lasted only a brief period, say three hours in the middle of the day, did that effect ridership more or less than a lingering drizzle that lasted all day but had less actual inches accumulated.

Likewise in test 2, the effect of the temperature on ridership. Using the average of the day is a little misleading on days that begin cold and end up much warmer. Are there trends throughout the day in the number of rides when the weather fluctuates? Overall there is a trend in weather and
season. Annual riders seem to be more committed to riding regardless of the season or the temperature, or even the precipitation. Purchasing an annual pass clearly makes that seem an obvious conclusion, but I was surprised by the significant drop in daily riders. Last winter was remarkably cold and that clearly had an effect on ridership. This year’s warmer fall, and hopefully winter, could alter the trend.

A. Further research

With this rich dataset, there is the potential for much more in-depth analysis. With my background in geographic information systems, I would be interested to see how geography plays into ridership. Is the distance traveled less on days that rain? Do you see a difference in rider type based on which stations are being used? In other words, are subscribers more likely to be moving in the manner of a commuter (exterior to interior) as opposed to daily customers who might be tourist sight-seers? I could also envision bringing in public transit data and seeing if there is a change in ridership of those venues on days when the weather seems less than ideal for biking.

In addition to the geographic information, for further analysis I believe it would be interesting to look at how ridership changes in relation to events in the city. Is there more ridership on days when there are events like the Air & Water Show or Lollapalooza? I would also look at more discreet datasets that target rain and snow as opposed to the large groupings of days where there was no precipitation present.

Finally, although I had the data, I did not go as in-depth in the annual rider demographics as I would have liked. I would be very interested in seeing how gender and age correlate to the weather in determining who rides. Of course, the data set is limited in that it does not include any demographics on daily riders, as this is not required information for renting a Divvy.

Using Hadoop and Cloudera’s advantage of handling large datasets, there are added potentials. I could envision Divvy linking not just weather data, but also CTA information for trains and buses. Imagine looking to see if there are any trends in Divvy ridership on days when there is heavy car traffic and delayed buses. Are people more likely to grab a Divvy to get to work if they think they’ll be late? This could also explain how riders use Divvy as a multi-modal transit system. They might ride Divvy from home to the CTA, take the train, and then Divvy to the office. Divvy could look at these trends and target similar commuters or offer daily specials to entice individuals to break out of their commuting habits. Cloudera/Hadoop could process this large quantity of information and make it more digestible.

REFERENCES

Fig. 1 Hadoop Ecosystem data flow

Fig. 2 Test 1 Rider type by rain

Fig. 3 Test 2 Average daily temperature by rider type

Fig. 4 Count of riders on days with an average temperature of 70.52°F

Fig. 5 Test 2 Ridership by temperature and season

Fig. 6 Test 2 Ridership by temperature and season, grouped by day

Fig. 7 Test 3 Average length of ride in minutes, by temperature
Fig. 8 Test 3 Average trip duration by user type and precipitation

Fig. 9 Test 3 Average trip duration by rider type and temperature