

Divvy and the Rain

Questions: How do different types of Divvy users respond to inclement weather conditions, in particular rain? What aspects of inclement weather have the greatest impact on ridership? Do different types of cyclists respond differently to different aspects of inclement weather.

Summary of Conclusions: Divvy users take fewer rides when it is raining out. Divvy users take shorter trips in terms of time when it is raining. Customers are much more sensitive to inclement weather relative to annual pass holders. Annual pass holders will in some cases be sensitive to rain conditions, and this response differs by gender. Male subscribers take slightly fewer trips based on how much it is presently raining while female subscribers take fewer trips if it is raining at all.

Data Sources:

- Divvy Data Challenge data sets provided by Alta Bikes
- Cycling Distances between stations downloaded from the Divvybrags github page
- Hourly Chicago weather data downloaded from Weather Underground

Methodology: The Divvy data was downloaded and station information was merged onto the trip data. Distances were easily merged onto this data set and subsequently simple data cleaning steps were performed. Hourly Chicago weather data was downloaded from the Weather Underground API and was merged onto the compiled data. The four most important variables added from this data set were precip, d_precip, rain, and d_rain. Their descriptions are below:

- precip: amount of precipitation in the last hour, in inches. Median value among non-zero values is only .02 inches
- d_precip: amount of precipitation that day, in inches. Median value among non-zero values is only .09 inches
- rain: binary indicating whether it has rained in the last hour. On average it rains in 4.6% of all hours
- d_rain: binary indicating whether it rains that day. On average it rains in 34.5% of all days

N.B. There is not a perfect correspondence between the two sets of variables, e.g. precip will be non-zero but rain will be zero. I believe this is primarily a function of the fact that many times when there is non-zero precipitation it is a negligible amount. The average precipitation in hours with positive precipitation and a "0" for the rain variable is .001 inches.

Subsequent to compiling the data, a series of regressions were run using the above four variables as the explanatory variables and the following four variables as the variable of interest:

- Total Rides: total number of rides taken in a given hour
- Average Ride Distance

- Average Ride Duration
- Average Ride Speed

These regressions were run on the following combinations of Divvy users.

- All Users
- Customers Only (24 Hour Pass Holders)
- Subscribers (Annual Pass Holders)
- Females Only (Excludes Customers)
- Males Only (Excludes Customers)

Gender data was not available for non-subscribers.

For the purposes of this report variables were typically only considered significant if they had a p-value of .05 or lower.

Conclusions: The table on the next page summarizes all key conclusions taken from this analysis. The most significant conclusions, however, are summarized below.

- Customers respond very strongly to the “rain” and “d_rain” variables. On a sunny day customers will take an average of 84 rides an hour, but on rainy days that number drops by nearly 21 rides an hour. If it is currently raining that figure drops an additional 32 rides, so in hours when it is raining customer ridership drops by more than 60%
- Subscribers, on the other hand, seem largely unfazed by inclement weather. On sunny days there are on average 91 subscriber trips taken every hour. This number drops slightly in response to current precipitation levels (on the order of a few riders), but otherwise does not drop at all in response to inclement weather.
- Male subscribers only respond statistically to the current precipitation levels. In the median rain storm this would amount to hardly any change at all
- Female subscribers only account for 22 rides per hour on average, but may lose as many as 4 rides in a given hour if it rains that day. Otherwise they seem largely insensitive to weather conditions.
- On a whole trip distance seems to shrink if it is raining.
- For all groups rides take less time when it is raining
- Overall riders ride slightly faster in response to rain and d_rain. This could either be riders struggling to get out of the rain, or a result of the fact that only more serious riders will be riding on rainy days

This table contains a brief summary of all regression results. Rows contain the name of the population subset categories, while the columns contain the variable of interest. In all cases the four variables used were those described above. In all case data was limited to trips with an average speed of less than 25 mph, a length of less than 2 hours, and a distance greater than 0 (to control for trips returning to the same station).

Summary of Regression Results by Poulation Category and Variable of Interest

	Total Rides	Average Ride Distance	Average Ride Duration	Average Ride Speed
All Rides	Significant drops in response to d_rain and rain	Odd results, slightly positive response to how much it is raining but strongly negative response if it is currently raining	Trips are almost 3 minutes shorter in rain, and about 30 seconds shorter if it rained that day	Speeds up by .1mph in response to rain and .4mph in response to d_rain
Customers (24 Hr. Passes) Only	Large drop in number of rides if it is raining or rained that day (around 30 and 20 rides out of an average 84 an hour)	Trips are .12 miles shorter in response to rain	Trips are .6 minutes shorter in response to d_rain, an additional 2.1 minutes shorter in response to rain	No significant change
Subscribers (Annual Passes) Only	No significant change	Trips are .12 miles shorter in response to rain	Trips are 1.4 minutes shorter in response to rain	No significant change
Females Only (Excludes Customers)	No significant change	Shorter rides taken if it is currently raining, regardless of how much	Rides are more than 1.5 minutes shorter if it's currently raining, and are slightly shorter based on the amount of rain that day	.1 mph increase in response to d_rain
Males Only (Excludes Customers)	No response to whether it is raining or rains at all that day, but large response at 5% level to how much it has rained that hour	Trips are .12 miles shorter in response to rain	Average sunny rides are 10 minutes long. They drop by 1.3 minutes in response to d_rain	No significant change

Challenges Encountered: the greatest challenge to performing this analysis was the tedious process of finding and analyzing reliable sources of hourly weather data. The data set presented here is actually the third source of hourly weather data I considered. The first one was the information provided by the OpenWeatherMap.org API which is a wonderful resource for up to the minute weather information provided by both reputable and amateur weather stations. This information can be queried directly or there is a dedicated python library for pulling information from the database. Unfortunately, after investing the time in learning this API, I found the data to be substandard in quality and that there were large gaps or inconsistencies in its quality.

Subsequently I found a more academic weather data source that provided direct access to Midway and O'hare weather data. After downloading and analyzing this data I found once again that it seemed at times inconsistent and that due to the json format in which it was stored, often times expected members of a json object didn't exist. I didn't want to write too much dynamic code given the quality of the data so once again I was forced to start from square one.

The third and final source of weather data I considered was the Weather Underground API. This data was consistent in format and fairly consistent in terms of quality. I still had to write "sleep" commands into my code, however, as the API limited a free account to only a few API calls a minute. Once downloaded it took a great while to compile the days into a single data source with the variables of interest.

Extensions: Once the data was compiled, the regressions themselves were fairly easy to run, but they are admittedly simplistic in nature. I believe that some of my results are interesting and perhaps even useful from a business perspective. Visually speaking, I would have liked to come up with a way to present my findings in a more aesthetically pleasing manner. Perhaps I could provide a means for user input with key weather events and then have some dynamic columns show the predicted hourly and daily traffic by various user categories.