Predicting BCycle behavior with weather

# Problem Statement

BCycle is a public bicycle sharing company owned and operated by the city of Austin since 2013. However, it has seen a sharp decrease in rides in the past year, particularly with UT students and in the downtown area. This is largely the result of the introduction of dockless scooters, operated by companies such as Lime and Bird, to Austin. This is of particular concern to Austin, as the city partnered with a non-profit to provide the BCycle service, and in fact used federal grants to purchase the roughly $2.5 million in equipment that makes up the BCycle system. As such, declining usage of BCycle means that Austin’s large investment is going to waste.

We hope to be able to use data analytics to preserve the value of Austin’s investment in BCycle. For instance, as any bicycle owner has noticed, bicycles rust and lose value when exposed to rain. It pays, then, to understand user behavior dependent on the weather.

Furthermore, using the BCycle data, and weather data, can we identify why customers use BCycle? For instance, how long do their rides last and where do their rides go? Is this dependent on the day of the week or the weather? Can we predict this behavior? Do riders use BCycle in different ways, i.e. can we identify distinct and different types of BCycle users?

We are also interested

What is the problem? How did you convert the problem into a data science problem? Why does the problem matter?

# Solution

### Data used: We examined two datasets. The first, primary dataset details B-Cycle usage in Austin from 21 December 2013 to 31 October 2018 inclusive. It contains 1.08 million rows—with each row detailing a single BCycle—and twelve features. Features include: date, check-out time, trip duration, check-out and check-in location, and membership type of the user. The second, supplementary dataset details weather in Austin from 21 December 2013 to 31 July 2017 inclusive. It has 1319 rows—one for each day—and 21 features, including date; high, low, and average temperature, humidity, wind speed, visibility and dew point; precipitation in inches; and any weather events that occurred.

### Approach and assumptions

Include data science approaches, nuances, research, and any innovations in procedure used.

### Application of Darwin

Why was Darwin used? How was Darwin used? How was the team’s experience with Darwin—i.e., what features of Darwin were helpful, and what features did you wish Darwin had?

# Team Engagement

How did the team divide roles? How was team participation?

# General Challenges

Where did the team run into challenges? E.g., research, machine learning, data, tooling, coordination, etc. What worked well, what was more difficult than anticipated, and what did you try without success?

# Next Steps

What could be done with this solution next?