Datetime & Availability:

1. Is data available for all days and time periods?

A green and white grid

Description automatically generated

* Above is the availability chart

A calendar with blue and green squares

Description automatically generated with medium confidence A blue and black table

Description automatically generated with medium confidence

* High number of Tuesday and Wednesday data is available.

1. Which is the busiest time/date/day of the week?

Busiest time is Morning and afternoon, followed by Evenings   
A screenshot of a computer

Description automatically generated

Station Information:

1. The top 5 popular/ busiest stations:
   1. Stations where cycles are removed quickly are very popular. (Less average dock\_wait\_time)

A screenshot of a computer

Description automatically generated

Do not use the above graph- as it captures in which station the cycles are removed immediately, but we need where more # bikes are docked and undocked rather how fast they are done.

* 1. Stations where bikes are continuously removed and docked (Highly changing bike and dock available numbers)

A graph with blue and white text

Description automatically generated with medium confidence

A map of a city

Description automatically generated

* Able to understand that stations around the university is the most popular- especially one next to Academic Building, Rec center, C4C, Norlin, Rec center and bus stand.
* But why is **13th & College** and **Timber Ridge** popular?

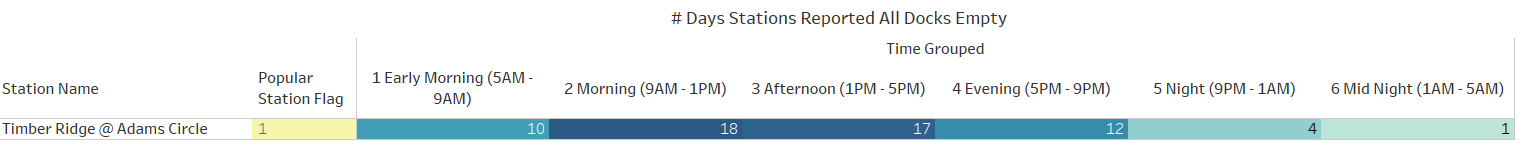
1. Do the busiest/ popular stations have high capacity?

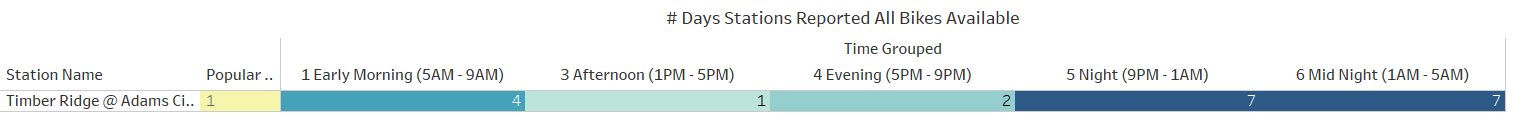
A screenshot of a graph

Description automatically generated

* Half of the popular stations have high capacity. Or this could be vice-versa.
* Still, why is 13th & College and Timber Ridge popular?

1. Why are these stations busy? Does it correlate with any of the features?
   1. Why is Timber Ridge popular?





* Timber Ridge station reported full of bikes in the non-class times and reported empty during class times. This means that the **station could be near college students’ residential area.**
  1. Why is 13th & College popular?

1. Stations that mostly have All docks empty-
   1. Norlin station? It is non-busy after evenings, because it have all docks empty after evenings?

* Yes, it is common that all docks are empty in the evenings, but this is a common trend among all popular stations
* However, Norlin is exceptionally reported All Bikes Available in the mornings more number of times than any other- **meaning people just use this station in the mornings more and do not use in the evenings**.

Usage & Wait Time:

1. What is influencing the bike usage? (Day, time, temperature, classes, etc):
   1. Correlation between factors vs sum (new docks available)
   2. Snake plot between factors
2. What is the variance of average wait times between stations?

Temperature Influence:

1. Influence of temperature in cycle usage?

Others

1. Dataset summary? Which columns are useful and redundant?
2. What is the ideal timeframe required to record the data?
3. Where to update the dock from 2.0 to 3.0? Is it needed to update them?

Final Dashboard Visualization:

1. Total number of stations? How many of them are currently working?

Check the distribution of wait times across various stations- divided by factors.:

* Filter only Morning-Evening?
* Plot quantile plots?
* Popular vs unpopular stations

A screenshot of a graph

Description automatically generated

* + Popular station have an average wait time of 91 mins however, it is only around 30mins in the busy times of the day.
* Across days for popular stations:

A screenshot of a computer

Description automatically generated

* + The wait times are lower in the weekdays than weekends in the popular stations:
    - For example, the wait time is only 12.9 mins Friday Morning.
    - Exceptions: Friday night has lesser wait times than other nights- Partying?

A screenshot of a computer

Description automatically generated

* + C4C is has the lowest waiting time- hence the populist station.
* Variance across stations, time and days?

A screen shot of a graph

Description automatically generated

* + Variance is low during **regular class days.**
  + Variance is low between **morning to evening**.
  + Variance is low in the **C4C station**.
* Same weekdays have similar distributions?
* Across different weekdays? Example, variance is more on Monday than Sunday?
* Mean is more on cold days?

Learn and think how to incorporate Expo in the data.

Go through Santhosh’s modeling code

Use learnings from EDA in Santhosh modeling:

* Bucket times? Take only morning – evening?
* Only popular stations?