Data Cleaning Notes

1. Cleaning steps:

* Downloaded 12 months of data from Dec 23 – Nov 24
* Made copies of monthly data sheets and put them all in one workbook in excel
* Formatting datetimes to all be the same in each sheet (m/d/yyyy h:mm)
* Created two new columns in each sheet, ride\_length and day\_of\_week and formatted the times to 37:30:55 and the days to numbers (Sun=1, Sat=7)

1. Data Notes:

* There are multiple sheets that are missing the start and end stations
* There are also sheets that are missing date stamps for the trips
* Not all sheets have the same datetime format
* Not all sheets have rounded off the long-lat data

1. R Analysis Notes:

* Analysis shows that the members typically ride the most from 5am – 10a, it dips then picks back up around 3p – 8p which would indicate these people are commuting to and from work during typical work hours
* The only stations that have more casual users than member users are stations located close to tourist locations:
  + Streeter Dr & Grand Ave
  + Shedd Aquarium
  + Dusable Harbor
  + Buckingham Fountain
  + Adler Planetarium
* A side-by-side comparison of the trip durations shows casuals ride longer and further suggesting trips to leisure destinations
* Side-by-side comparison of overall trips throughout the week show members clearly make up the majority of riders

1. Recommendations:

* Push ads in the app when casual riders attempt to use the service
* Install ads around the most frequently used casual rider stations (billboards, posters, etc.)