SYDNEY DATA DICTIONARY

==========

CSV 1 - Bicycle_survey_sites.csv
Dimensions: 110 x 5

DATA DICTIONARY
X - Longitude coordinates of Sydney. North is positive, South is negative
Y - Langitude coordinates of Sydney. North is positive, South is negative
OBJECTID - ID based on numerical record count
SiteID - Unique ID representing each intersection
Intersection - The unique location where intersections cross

Method
- Used count unique to check the number of entries in each column (confirmed 109 unique entries each)
- Number of entries in each column matched the number of intersections
- checked the long/lat coordinates of sydney and found similar numbers.
- Assuming that positive means North and East, whereas Negative means South and West- -
- Used filter to sense-check the data

Note on data formats:
- X, and Y coordinates have up to 5 decimal places but can be inconsistent.
- ObjectID and SiteID are numeric
- Intersections are all string.

No statistical calcualtions done as it seemed irrelevant to the dataset.
==========

CSV 2 - Bicycle_count_surveys
Dimensions 2216x11

Data Dictionary

SiteID-Unique Site ID Corresponding to intersection. Primary Key used to join to CSV1.
Month-Month where the number of cyclists were tracked
Year-Year where the number of cyclists were tracked

TotalCount-Total Number of cylcists summed from 6-9am and 4-7pm
ObjectId2-Record Id of this dataset?
Time_0600-Number of cyclists between 6pm-7pm
Time_0700-Number of cyclists between 7pm-8pm
Time_0800-Number of cyclists between 8pm-9pm
Time_1600-Number of cyclists between 4pm-5pm
Time_1700-Number of cyclists between 5pm-6pm
Time_1800-Number of cyclists between 6pm-7pm

Anomalies

29 records do not have an equal sum with the total count.
Object IDs are not consecutive.

Method:
- Utilised Excel and VLOOKUP to join both data sources via SiteID
- Filtered each column to find Unique values

Observations
- March 2018 is missing data, as the website says
- Some total counts do not match
- the ObjectId2s are not consecutive
- Majority of the data points have outliers above the upper limit, indicating right skew