

Concordia University

Gina Cody School of Engineering and Computer Science

Department of Building, Civil and Environmental Engineering

CIVI 691K: Big Data Analytics for Smart City Infrastructure

Winter 2019

Instructor

MAZDAK NIK-BAKHT

Homework assignment #1

Submitted by,

Vivekkumar Chatrola- **40059267**

1. The Context–

This assignment is prepared using year 2016 dataset of BIXI Montreal station code 6184. Station de vélo Bixi 6184 is located at Le Plateau-Mont-Royal, Montréal, QC H2J 1W2. (45°31'28.8"N 73°34'57.5"W; Latitude: 45.524664, Longitude: -73.582654). It is on the Rivard st near its intersection with Mount-Royal Ave E street. STM Mount-Royal Metro station (less than 1m) and STM Laurier Metro station (450m – 6 minutes' walk) are the closest sunbways. There are nearby bustsops for following STM buses: 11, 29, 30, 97, 361, 368, 427, 711. Café Myriade - Le Plateau, La Matraca,

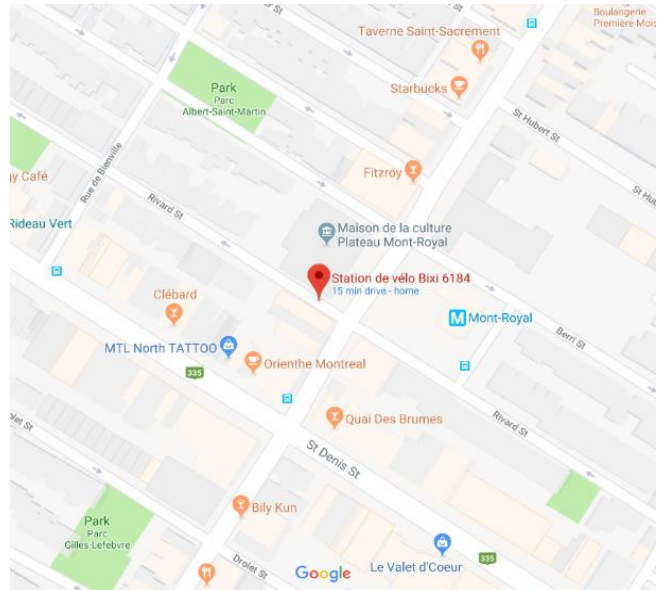


Figure 1 Google map screen shot of bixi station 6184

A&W Canada, Starbucks are the nearest restaurants. Maison de la culture Plateau Mont-Royal, PJC Jean Coutu, Library of Plateau-Mont-Royal, Parc Albert-Saint-Martin, Montreal Conservatory of Music, Banque Nationale, BMO Banque de Montréal, Théâtre du Rideau Vert are nearest attraction. There are total **75219 examples and 24 attributes** in the bixi station 6184, 2016 dataset.

2. Understanding the Attributes –

- In this dataset, there is no missing values and inconsistent values.

Table 1 Understanding the attributes

| Attribute name | Data type | Unit | Center tendency statistics | Spread Statistics |
|--------------------|-------------|-----------|---|---|
| day | Polynominal | | Mode- Thursday (11434) | |
| start_time_stamp | Integer | second | Mean- 24493507.451 | Deviation- 80760.990 Range- (24344891, 24654523) |
| start_date | Polynominal | Date_time | Mode-2016-07-31 15:01 (8) | |
| start_station_code | Polynominal | | Mode- 6184 (39711) | |
| start_station_name | Polynominal | | Mode- Metro Mont-Royal (Rivard / du Mont-Royal) (39711) | |

Homework Assignment #1
Date: Sat, Feb 9th, 2019

CIVI 691
Chatrola, Vivekkumar

| | | | | |
|-------------------------|------------|-----------|---|---|
| start_station_latitude | Real | | Mean – 45.525 | Range- (45.444, 45.577) |
| start_station_longitude | Real | | Mean- -73.582 | Range – (-73.671, -73.509) |
| end_date | Date_time | Date_time | Mode- 2016-10-01 (11) | |
| end_station_code | Polynomial | | Mode- 6184 (35946) | |
| end_station_name | Polynomial | | Mode- Metro Mont-Royal (Rivard / du Mont-Royal) (35946) | |
| end_station_latitude | Real | | Mean- 45.525 | Deviation- 0.009 Range- (45.434, 45.581) |
| end_station_longitude | Real | | Mean- -73.581 | Deviation- 0.011 Range- (-73.671, -73.495) |
| duration_sec | Integer | seconds | Mean- 644.662 | Deviation- 533.342 Range- (61, 7196) |
| is_member | Bionominal | | Mode- True (62772) | |
| start_weather_type | Polynomial | | Mode- sky is clear (17285) | |
| start_humidity | Integer | | Mean- 63.855 | Deviation- 18.886 Range- (16, 100) |
| start_wind_speed | Integer | | Mean- 3.495 | Deviation- 2.217 Range- (0, 14) |
| start_temperature | Real | Celsius | Mean- 17.697 | Deviation- 7.084 Range- (-4.076, 32.290) |
| start_air_pressure | Integer | | Mean- 1014.919 | Deviation- 8.086 Range- (990, 1039) |
| end_weather_type | Polynomial | | Mode- sky is clear (17213) | |
| end_humidity | Integer | | Mean- 63.665 | Deviation- 18.918 Range- (16, 100) |
| end_wind_speed | Integer | | Mean- 3.508 | Deviation- 2.216 Range- (0, 14) |
| end_temperature | Real | Celsius | Mean- 17.751 | Deviation- 7.082 Range- (-4.076, 32.290) |
| end_air_pressure | Integer | | Mean- 1014.909 | Deviation- 8.077 Range- (990, 1039) |

3. Understanding the Records–

Based on all the graphs, following insights should be concluded:

1. Thursday is the most favorite day for bixi users.
2. Member tends to travel more than non-members
3. Whenever there is a heavy rain, bixi use goes next to zero.
4. Marquette / du Mont-Royal is the most used destination from station 6184
5. Surprisingly, Sunday is the least busy day of the week.
6. Du Mount Royal/Clerk is the most used origin bixi station leading to station 6184

4. Correlational Analysis–

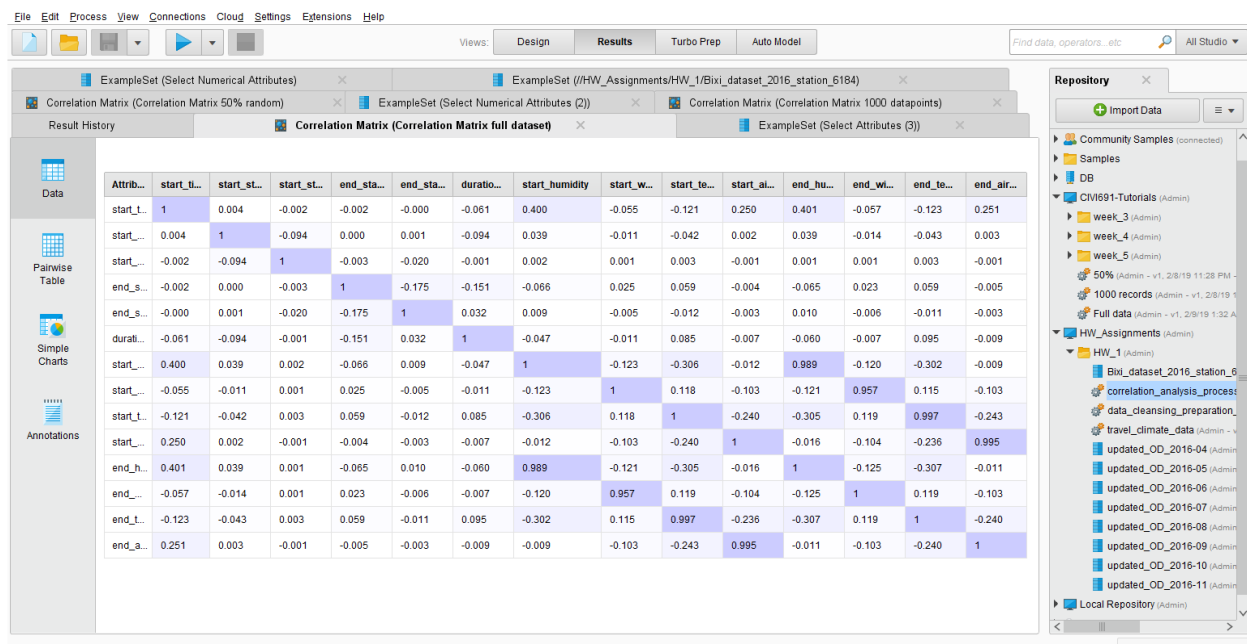


Figure 2 Correlation matrix for full dataset

Table 2 Correlation between pair of attributes

| Very strong correlation | Strong correlation | Some correlation |
|---|--------------------|---|
| Start humidity & end humidity (0.989) | | end humidity & start time stamp (0.401) |
| Start wind speed & end wind speed (0.957) | | start humidity & start time stamp (0.400) |
| Start temperature & end temperature (0.997) | | |
| Start air pressure & end air pressure (0.995) | | |

- There are no attributes with strong correlation in the dataset

5. Correlational Analysis for Reduced Data–

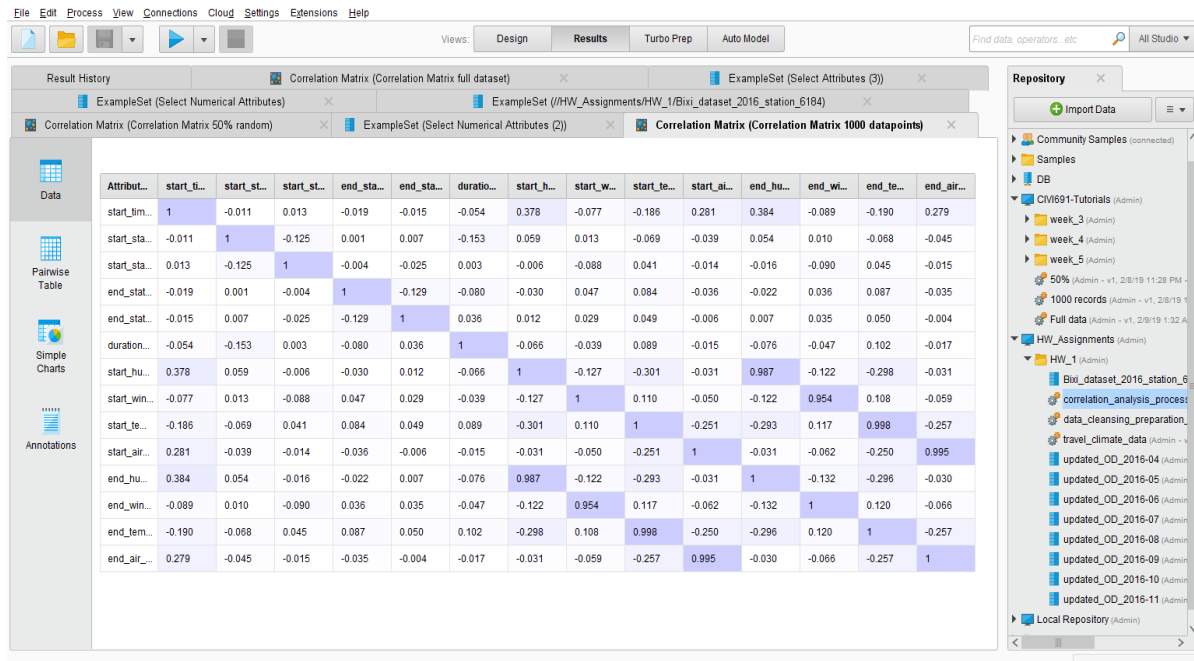


Figure 3 Correlation matrix 1000 records

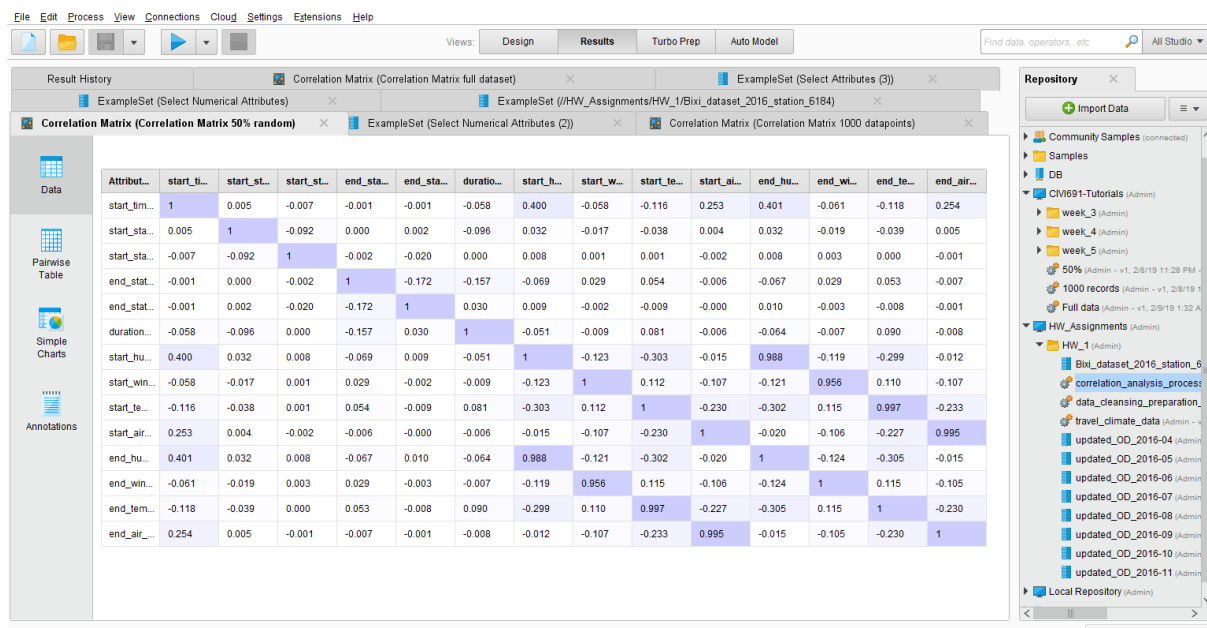


Figure 4 Correlation matrix 50% random datapoints

- All the three matrixes show similar correlation pattern because for the 1000 records correlation matrix we used similar distribution for weekdays by using role set operator & for the 50% randomly chosen data points, we used half of the dataset that's why it showed similar pattern.