Report on Edinburgh
Bike Count Data

Group members:
Sally Bai,
Yao Tong,
Xuanchen Yao

THE UNIVERSITY
of EDINBURGH

Table of Content

Introduction

Data Preparation and Processing

Data Analysis Methodology

Results on Detected trends

Conclusion

Introduction

1. Project Context
The City of Edinburgh Council (CEC) has brought value to improve public transportation in order to achieve a cycling and walking-friendly city, and more importantly, to enhance live quality. To approach people's cycling behaviour, path usage and traffic density, the data Bike counters were installed to record bike counts on an hourly-basis.
In addition to the data collection of CEC cycle team, the organization "Spookes" has also installed counters for promoting cycling related projects. Our work will contribute to previous existing datasets from CEC cycle team to give the hypothesis and compelling statistics.

2. Aims and Achievements
The main purpose for this project is to approach people's cycling habits in Edinburgh city - how did people perform in different time periods. Divided into different time session along the year 2015 and 2016, the datasets which provided by the city council of Edinburgh shows interesting findings by going through data visualisation.
There are a series of questions we aim to answer when dealing with data:
    Are there any irregular data worthy of exploring? What would be the potential cause?
    How different time periods effect on cycling traffic? Seasonality? Weekdays and weekends?
    Are there any relationships between bike counts and variables such as weather and altitude?

Data Preparation and Processing

1. Data Description
The data are from 19 off-road counters and 29 on-road counters from 2004 to 2016. Except the data has been collected constantly, temporarily missing data are due to battery failure which also shows an irregularity in graphs. Our project chose 13 sites out of 48 and restricted the period between 22/03/2015 and 18/06/2016.

There are total 5 columns, representing

counter_id: the data is collected via labelled automatic counter (1-48);
date: the day on which the data was collected (18/03/2010);
time: the data is collected on an hourly-basis (0~23);
channel_1: the direction of travel (north-bound);
channel_2: the direction of travel (south-bound)

2, Data Preparation
Each bike count site was collected in weeks. First step is to combine the site data according to their count id. Then the datasets were renamed according to their locations.

1. '25straitonpath'
2. '26stenhousedrive'
3. '29corstorphineroad'
4. '30dalryroad'
5. '34nicolsonstreet'
6. '36raeburnplace'
7. '40mayfieldroadnorthbound'
8. '41mayfieldroadsouthbound'
9. '42melvilledrivemain'
10. '43melvilledrivespur'
11. '44melvilledrivebus1southbound'
12. '46fishwivescauseway'
13. '47inverleithpark'

Data Analysis Methodology

All the data analysis and visualisation have implemented through ipython notebook. Additional libraries such as numpy and pandas have also been used. The format of datasets has also been altered to cater different time periods.
1. The traffic of channels is first added up. The change of bike counts can be presented in distribution plot, boxplot, regression plot and so on.
2. Other datasets: Relate weather data to cycling data with time, and probe the linear relationship
  https://www.ed.ac.uk/geosciences/weather-station/weather-station-data
3. Parameter of time period
the total traffic of each sites are added up to visualise the change within a day.
4. Approaching the altitude of each site's location with its traffic

Results on Detected trends
Plots shown below present various forms of clustering on bike counts. Time line is one of the axis to parse bike datasets.

1) Analysis on cycling traffic at different locations

The Amount of Bicycles Among 13 Locations (22/Mar/2015-18/Jun/2016)
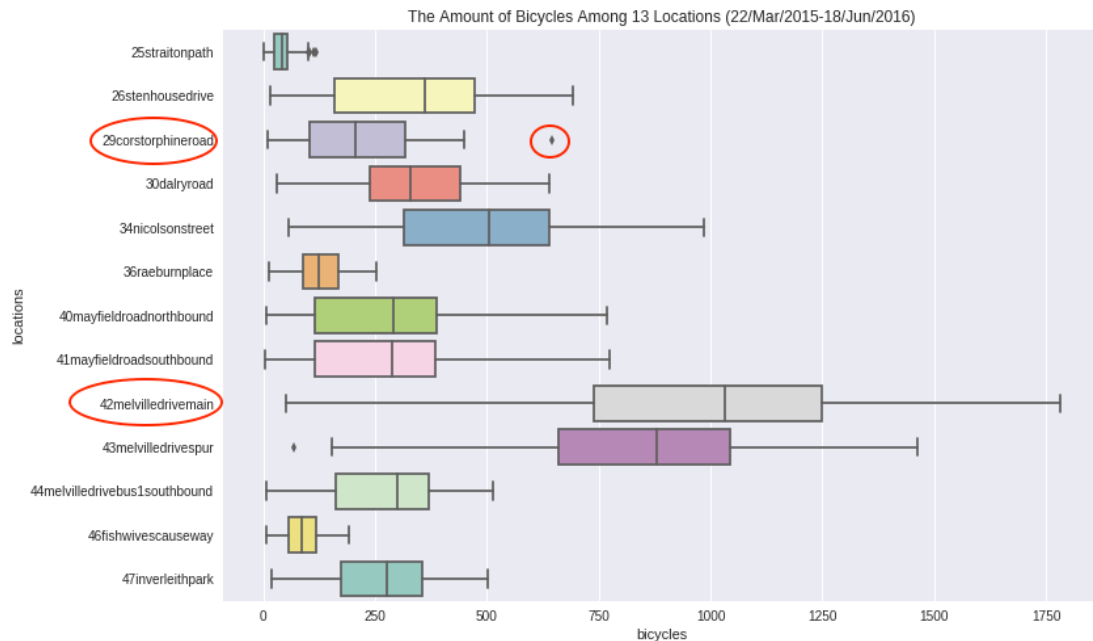
Comments:

The spread of traffic distribution has shown different statistic range. The interquartile of Site 42 and 43 are greater than other sites.

The strange outlier appears in counter 25

After searching online with the key words like '2015 Sep 6th' and 'bike', we found that there was an event happening right there then, which is just crossing the counter set at Corstorphine Road, namely, 29corstorphineroad.

Pedal for Scotland- Scotland's Biggest Bike Event – 5th, 6th, 13th September 2015. We can find the following link for further information about what happened during that time:

http://www.dailyrecord.co.uk/lifestyle/health-fitness/pedal-scotland-scotlands-biggest-bike-6103718

According to this findings, bike events can also affect the amount of bikes significantly.

2) The change of traffic from 04/2015 to 06/2016 (take counter 42/43 as example)

the change of the amount of bikes as time goes by
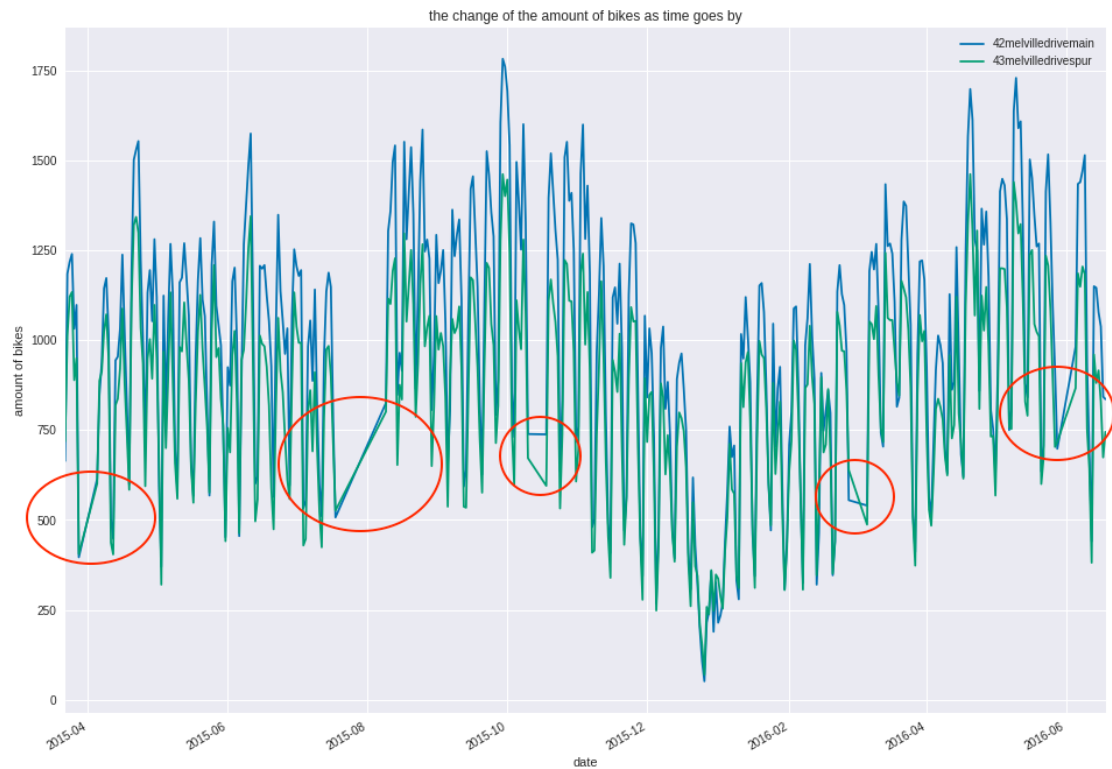
comments:
Unlike other fluctuations, the steady and straight increases appear in the red circles. The bike counters are considered out of service;

The bike counts over the year has fluctuated dramatically, and the peaked ups and downs indicated the different performance in days. In 2016/01 or so, the traffic hit its lowest point. To get a clearer point of view, we narrow down the time line:

the change of people biking behaviour during Christmas

data reached its bottom around Dec.26, which fit in line with the Christmas.

3) Approaching the performance of bike counts in days of a week
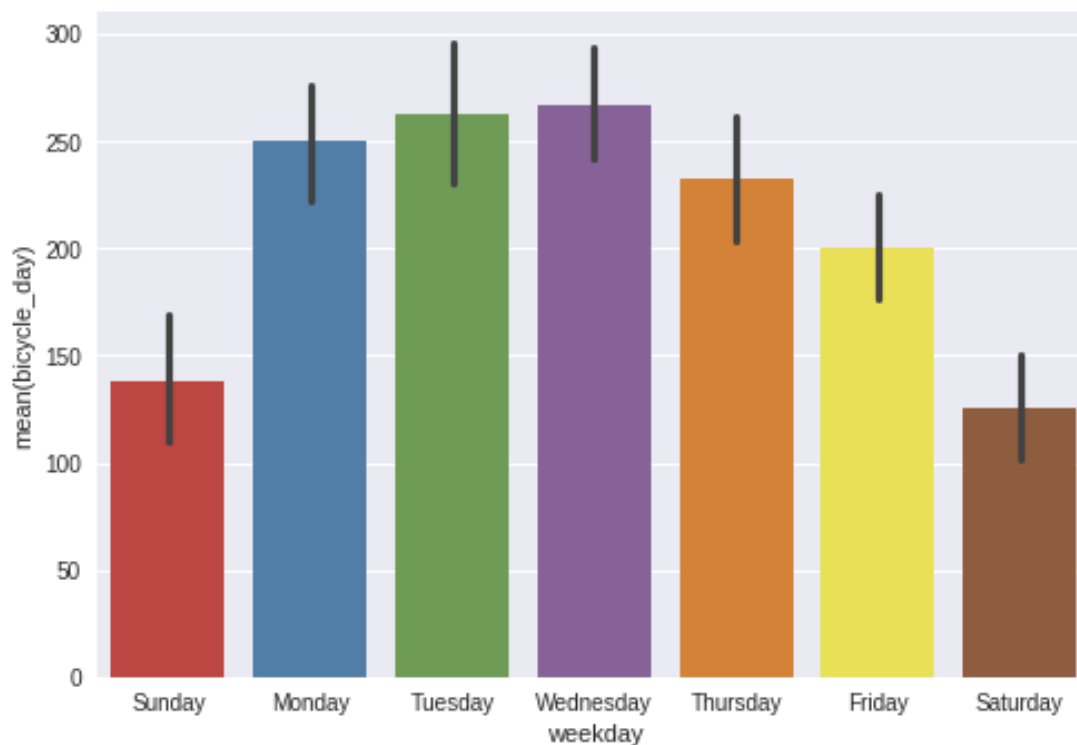
comments:
People prefer to go out for biking on weekday rather than weekends,
even taking confidence intervals into consideration

We get the hint of the decrease in weekends, which the determined factor of
bike traffic pertains to commuting. To peek into the traffic change within one
day, the boxplot below shows the dispersion of traffic in an hour basis:
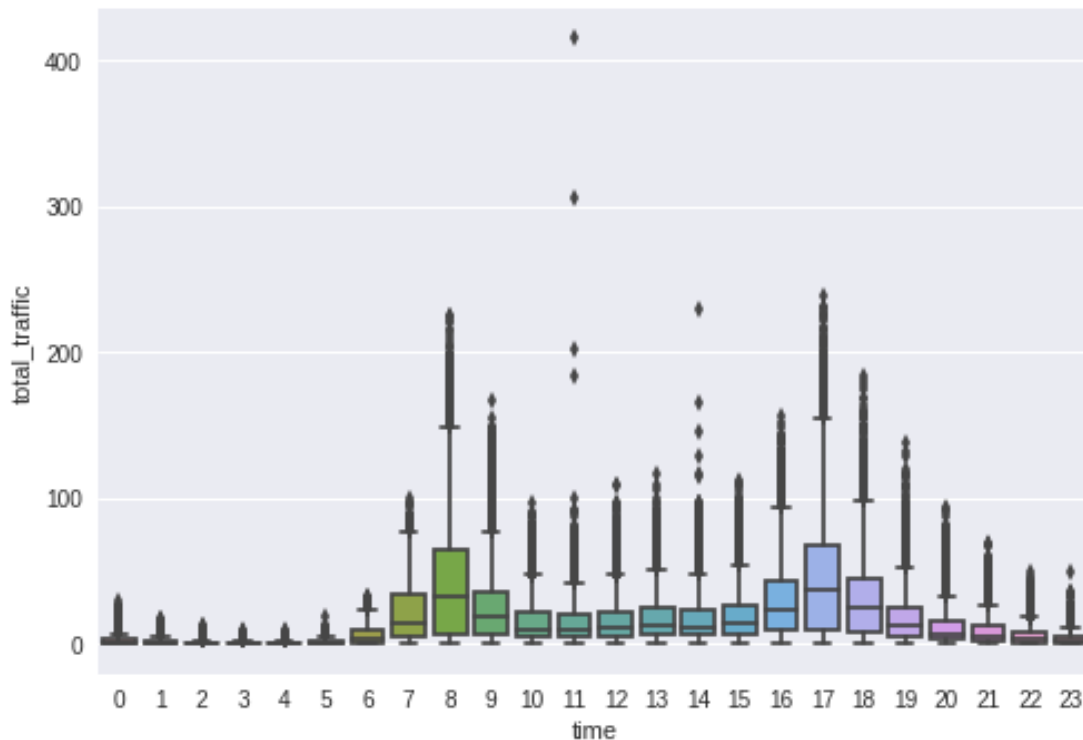
According to the graph, two rises occurs in 7 to 9 and 16 to 18. It proves that the main traffic in weekdays is from commuting.

4) Other exploratory variables

The correlations between weather and bicycle

The plot below shows a weak linear correlation.

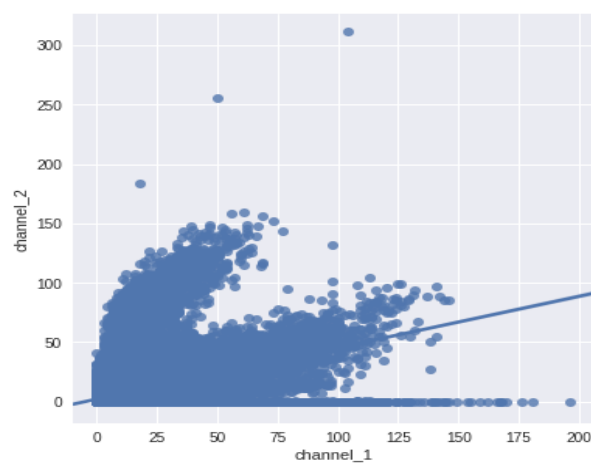In accordance with the it, the statistics is

rvaule = -0.23 -- Wind speed does have a weak negative relationship with the bike counts as rvalue is between -0.3 and 0.

pvalue = 0.001 -- We can accept our hypothesise as pvaule is less than 0.05.

intercept = 1137 -- When wind speed is 0, bike amount is likely to be 1137.
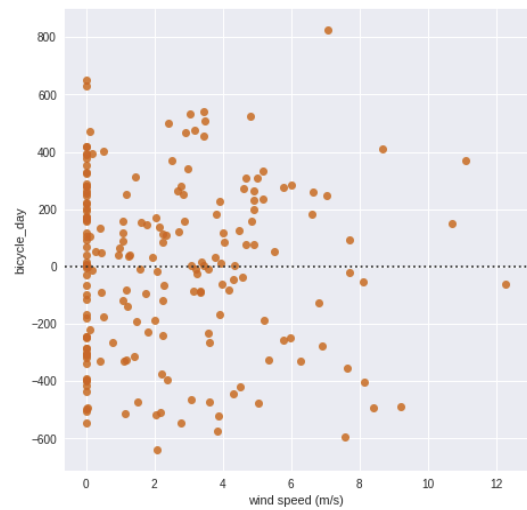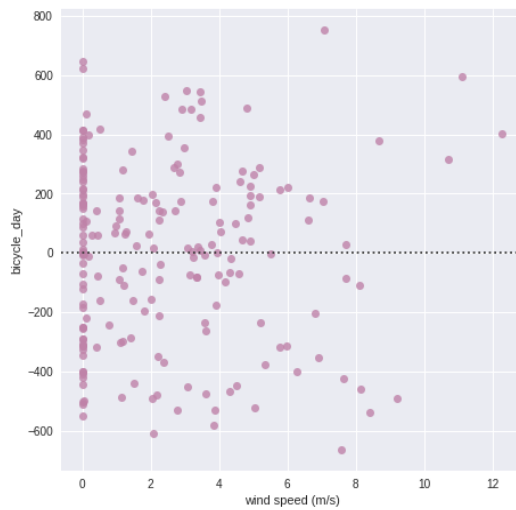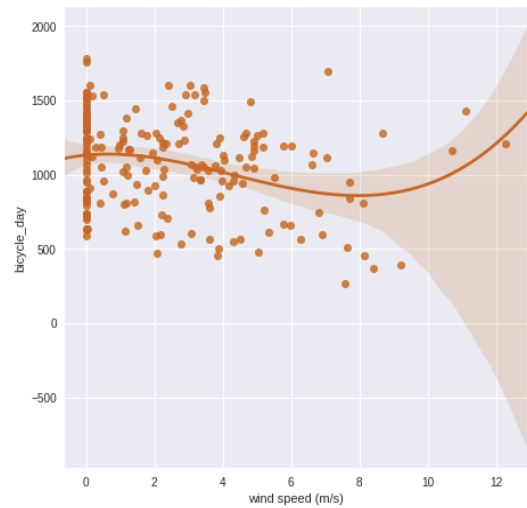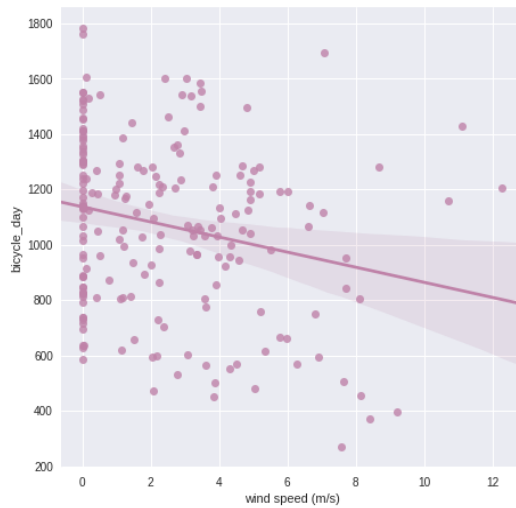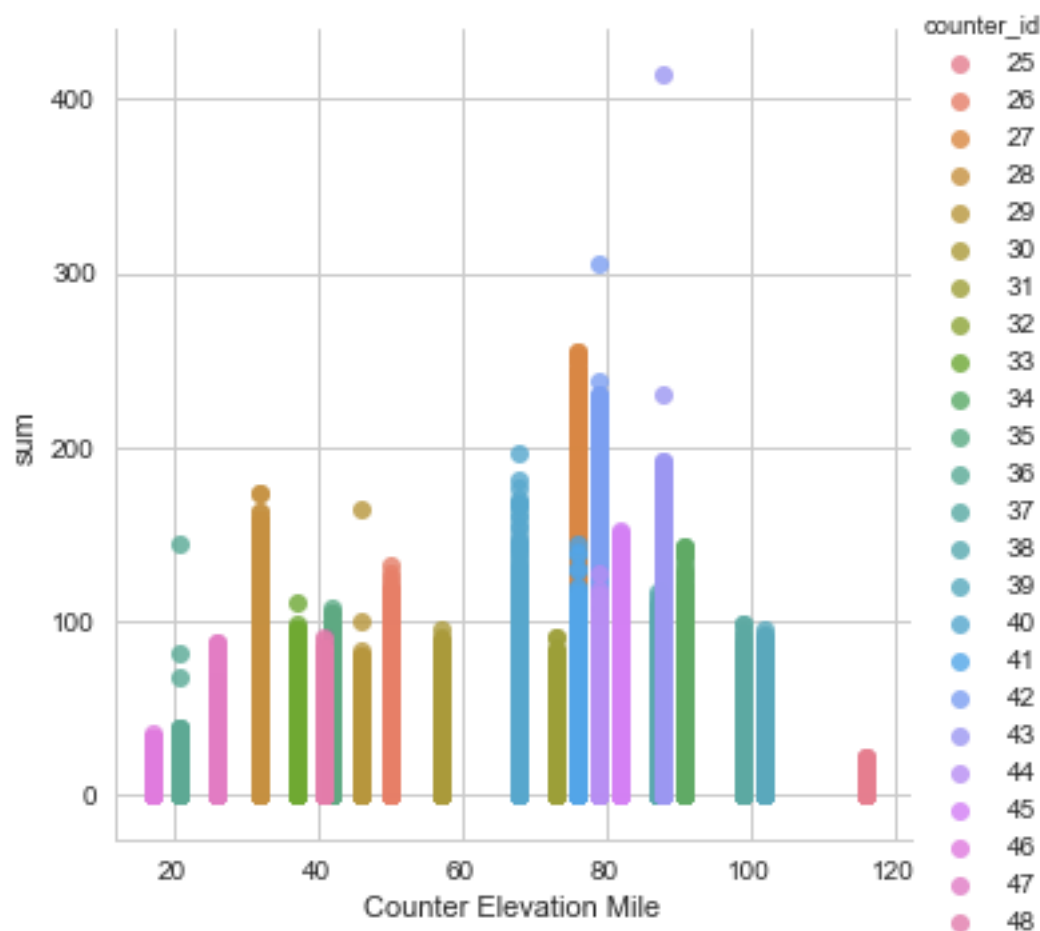
The relationship between two route directions

Comments:
Plot doesn't show a strong linear relationship. It indicates that people might choose different routes in a day.

The relationship between altitude and bike traffic

Conclusion

Our purpose for this project consisted in trying to approach the cycling status in Edinburgh by means of dataset analysis. We adopted the "Extract", "Transform", "Load" process to parse the cycling traffic over a number of time periods. Bike counts in weekdays are greater than in weekends, so people choose taking bike as commuting transportation. The specific factor such as bike campaign and Christmas day will rise and descend the bike traffic respectively.

Moreover, we focused on investigating the underlying pattern by combining other open datasets such as the weather dataset of city Edinburgh and altitude distribution. While the correlation coefficient of those variables is not as strong as chronic analysis, it revealed the potential features which stimulated more in-depth examination.