Assignment 3: Bicycle Counter Data

Shane McCarthy

shane.mc-carthy@ucdconnect.ie

26 April 2016

Part 1 – Calculate monthly summaries at each location

a. Methodology

The CRISP-DM data mining approach was used to help guide the data exploration journey, it consists of the following steps:

1. Business Understanding – establishing what the key objectives and requirements of the assignment are.
2. Data Understanding – developing an understanding of what the data means, evaluating the data quality, establishing how it links together and unravelling different naming conventions, formats and types.
3. Data Preparation – reshaping the data in preparation for modelling.
4. Modelling – experimenting with a series of models to determine which is the most suitable for our data.
5. Evaluation – evaluating how well the models fit our data using an out-of-time test sample.

(The sixed step of CRISP DM is "Deployment", this is outside the scope of the assignment)

Given the poor data quality of the assignment dataset a large proportion of time was spent on step two. To accelerate this data understanding process a function was developed that picked up all the column names in the dataset, gave each unique column a unique ID and then checked each sheet (each year) for every variable. The output of this function can be seen in the table below, variables are listed on the vertical with flags called "variable 2012…2015" listed along the horizontal, the proportion of observations missing is also captured.

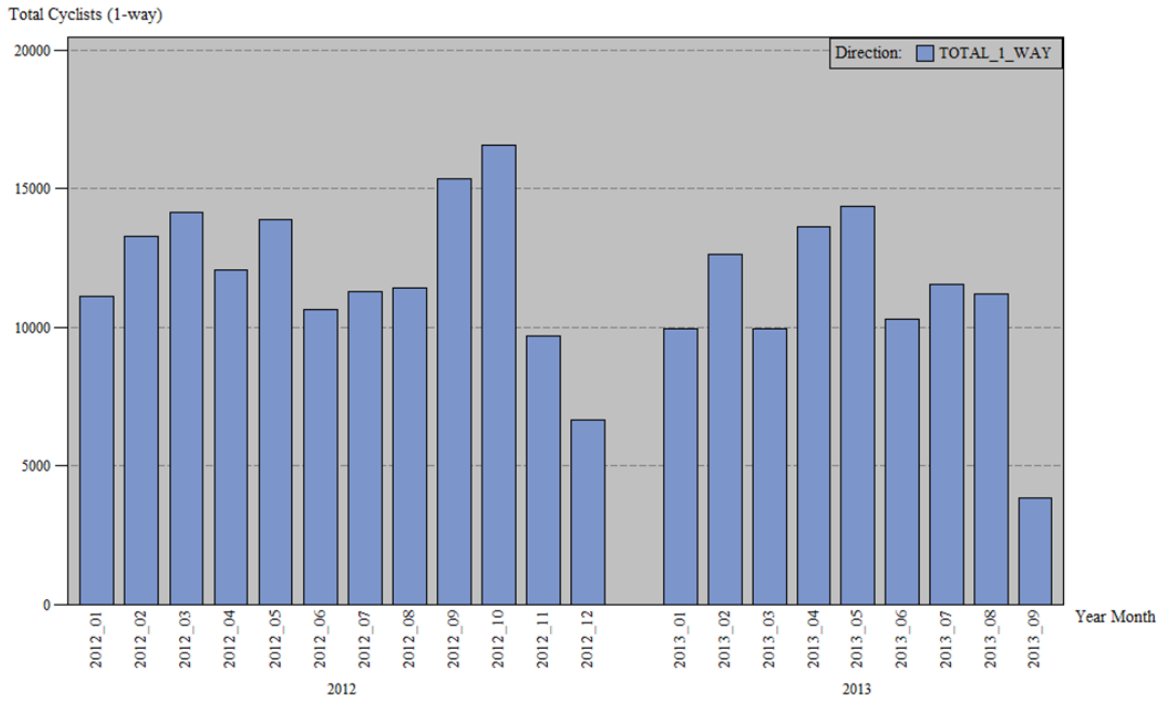
ID	variable	label	variable 2012	prop_missing 2012	variable 2013	prop_missing 2013	variable 2014	prop_missing 2014	variable 2015	prop_missing 2015
3	DATE	Date	1	0%	1	0%	1	0%	1	0%
16	TOTEM_N11_STILLORGAN_RD_	Location 1: Cyclists 1-way	1	0%	1	32%	0	100%	1	100%
1	BICYCLE_IN	Location 2: Cyclists in	1	0%	1	0%	1	0%	1	0%
2	BICYCLE_OUT	Location 2: Cyclists out	1	0%	1	0%	1	0%	1	0%
19	WALKING_IN	Location 2: Pedestrians in	1	0%	1	0%	1	0%	1	0%
20	WALKING_OUT	Location 2: Pedestrians out	1	0%	1	0%	1	0%	1	0%
4	GLENAGEARY	Location 2: Total of walking	1	0%	1	0%	1	0%	1	0%
7	N11_MONTROSE	Location 3: Total of cyclists	1	7%	1	0%	1	0%	1	0%
5	IN__TOWARDS_CITY_CENTRE	Location 3: cyclists in	1	7%	1	0%	1	0%	1	0%
8	OUT_AWAY_FROM_CITY_CENTRE	Location 3: cyclists out	1	7%	1	0%	1	0%	1	0%
12	ROCK_ROAD__BUS_LANE_BESIDE_	Location 4: Cyclists 1-way	1	7%	1	0%	1	0%	1	0%
10	ROCK_ROAD_PARK_IN	Location 5: Cyclists in	1	6%	1	0%	1	0%	1	0%
11	ROCK_ROAD_PARK_OUT	Location 5: Cyclists out	1	6%	1	0%	1	0%	1	0%
13	ROCK_ROAD__PARK	Location 5: Total cyclists	1	6%	1	0%	1	0%	1	0%
15	TOTEM_IN	Location 6: Cyclists in	0	100%	1	1%	1	47%	1	0%
17	TOTEM_OUT	Location 6: Cyclists out	0	100%	1	1%	1	47%	1	0%
14	TOTEM_CLONSKEAGH_ROAD	Location 6: Total of cyclists	0	100%	1	1%	1	47%	1	0%
18	TOTEM_ROCK_ROAD	Location 7: Cyclists 1-way	0	100%	1	1%	1	30%	1	100%
6	N11_ECO_TOTEM	Unkown	0	100%	0	100%	0	100%	1	94%
9	ROCK_ROAD_INBOUND	Unkown	0	100%	0	100%	0	100%	1	78%

Once the data was extracted from the assignment dataset, cleansed and merged it was passed into a function which aggregated it to a monthly level and outputted a histogram plot along with a table summary.

b. Results & Observations

This section includes summaries for each of the seven locations along with observations on data quality and trends.

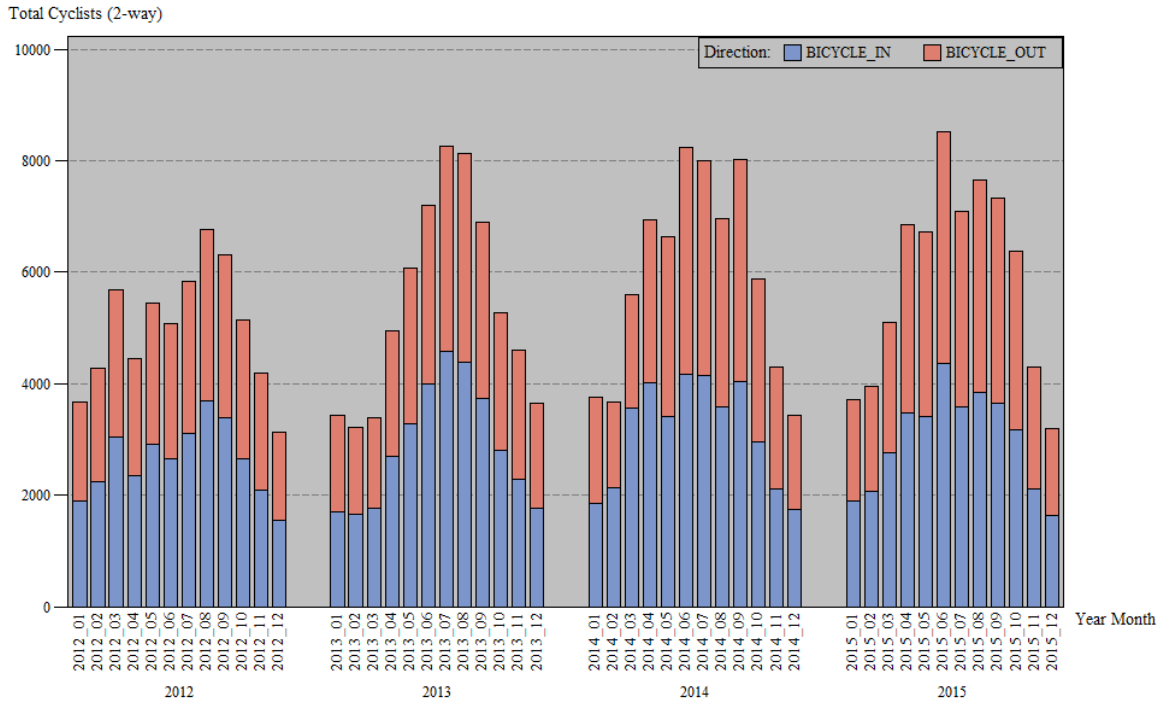
Location 1: N11 Stillorgan Road Totem (1-way)



	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	11,120	13,255	14,141	12,056	13,895	10,618	11,285	11,411	15,367	16,549	9,685	6,631
2013	9,944	12,615	9,949	13,626	14,356	10,291	11,545	11,197	3,832	0	0	0
2014	0	0	0	0	0	0	0	0	0	0	0	0
2015	0	0	0	0	0	0	0	0	0	0	0	0

Location 1 - N11 Stillorgan Road Totem is a “one-way” location so directional data is not available. It is clear from the plot and table that the data quality is poor for this location with missing values from Oct 2013 onwards.

Location 2: The Metals, Glenageary Dart Station

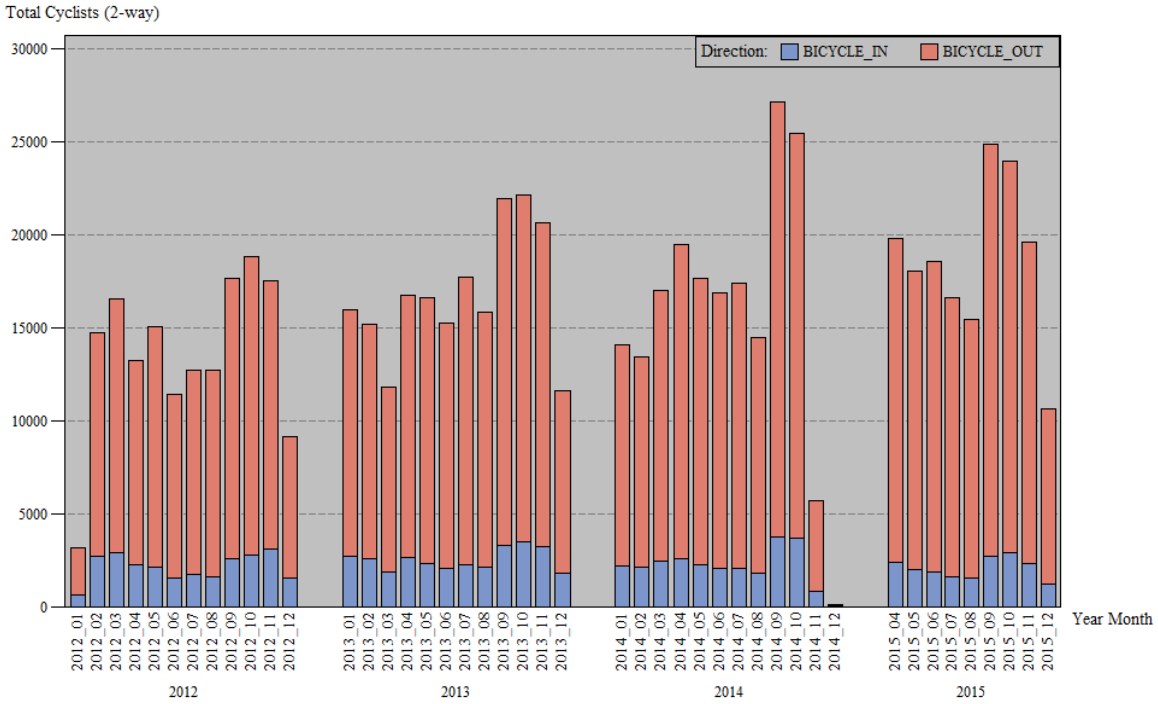


	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	3,680	4,268	5,684	4,442	5,456	5,073	5,829	6,764	6,303	5,138	4,201	3,138
2013	3,443	3,210	3,381	4,943	6,072	7,190	8,262	8,136	6,892	5,269	4,606	3,659
2014	3,764	3,671	5,593	6,936	6,641	8,229	8,010	6,952	8,025	5,880	4,298	3,432
2015	3,713	3,945	5,098	6,855	6,725	8,519	7,091	7,654	7,333	6,374	4,292	3,205

Location 2- The Metals, Glenageary Dart Station is a two-way location therefore a directional breakdown can be plotted. The data quility for this location is good with a clear seasonal trend being observed.

Note: the pedestrian counter data included in the assignment dataset has been excluded as the objective is to analyses cycle counter data only.

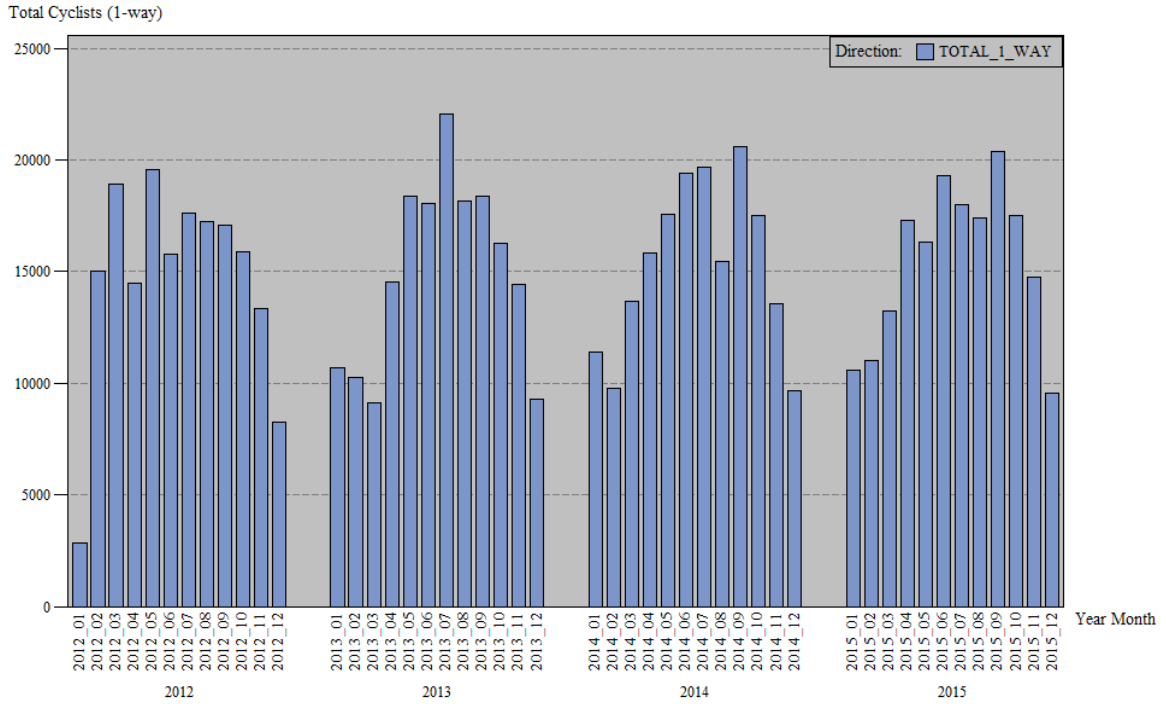
Location 3: N11 Montrose (2-way cycle track)



	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	3,184	14,728	16,525	13,192	15,033	11,397	12,692	12,724	17,638	18,774	17,503	9,158
2013	15,932	15,187	11,761	16,703	16,624	15,235	17,727	15,807	21,933	22,127	20,631	11,578
2014	14,077	13,420	16,987	19,430	17,642	16,873	17,385	14,444	27,143	25,417	5,700	51
2015	0	0	0	19,798	18,017	18,569	16,573	15,434	24,841	23,933	19,578	10,617

Location 3- N11 Montrose is again a two-way location, therefore a directional breakdown can be plotted. The data quality for this location is questionable, particularly for the inbound count which is dwarfed by the outbound count. One logical explanation for this is that the counter is installed on a one-way street with a small number of risk adverse cyclists accounting for the inbound count.

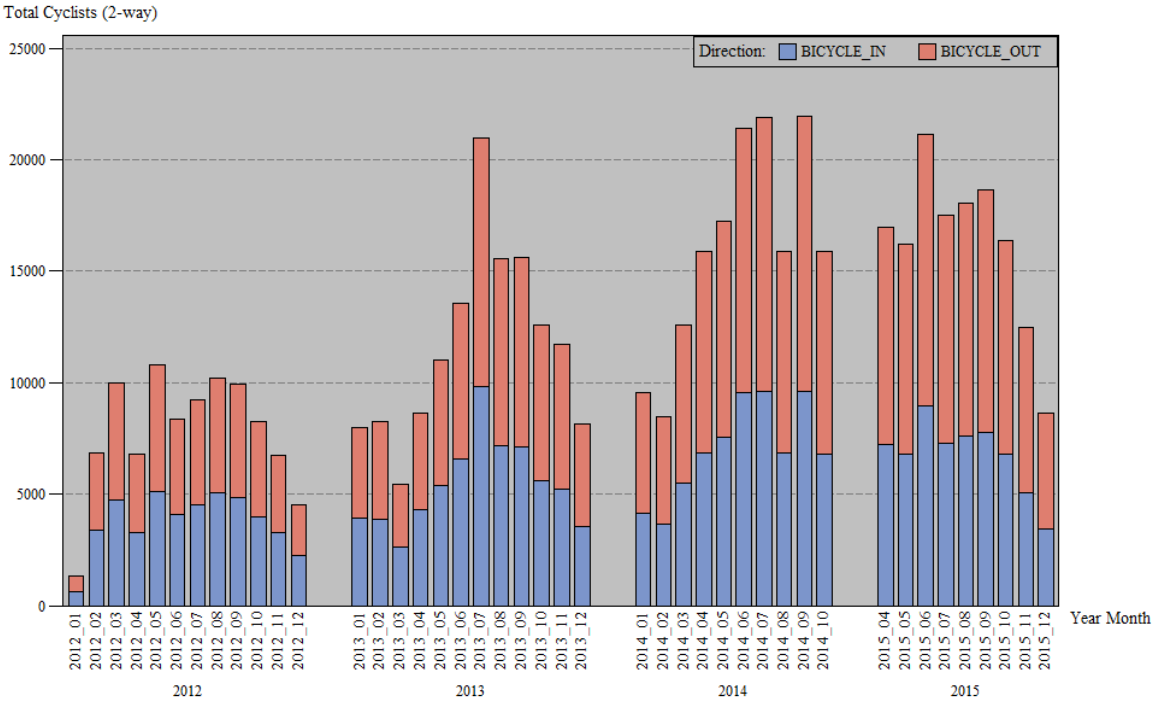
Location 4: Rock Road Bus lane (1-way)



	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	2,834	15,001	18,914	14,467	19,551	15,753	17,619	17,216	17,065	15,879	13,325	8,263
2013	10,693	10,277	9,122	14,560	18,376	18,049	22,075	18,145	18,401	16,279	14,409	9,294
2014	11,407	9,784	13,672	15,854	17,548	19,385	19,662	15,471	20,586	17,509	13,575	9,657
2015	10,571	11,037	13,229	17,317	16,344	19,292	18,005	17,415	20,375	17,531	14,756	9,574

Location 4 - Rock Road Bus lane is a “one-way” location so directional data is not available. With the exception of a number of outlier months in 2012 the data quality is relatively good with a clear seasonal trend being observed.

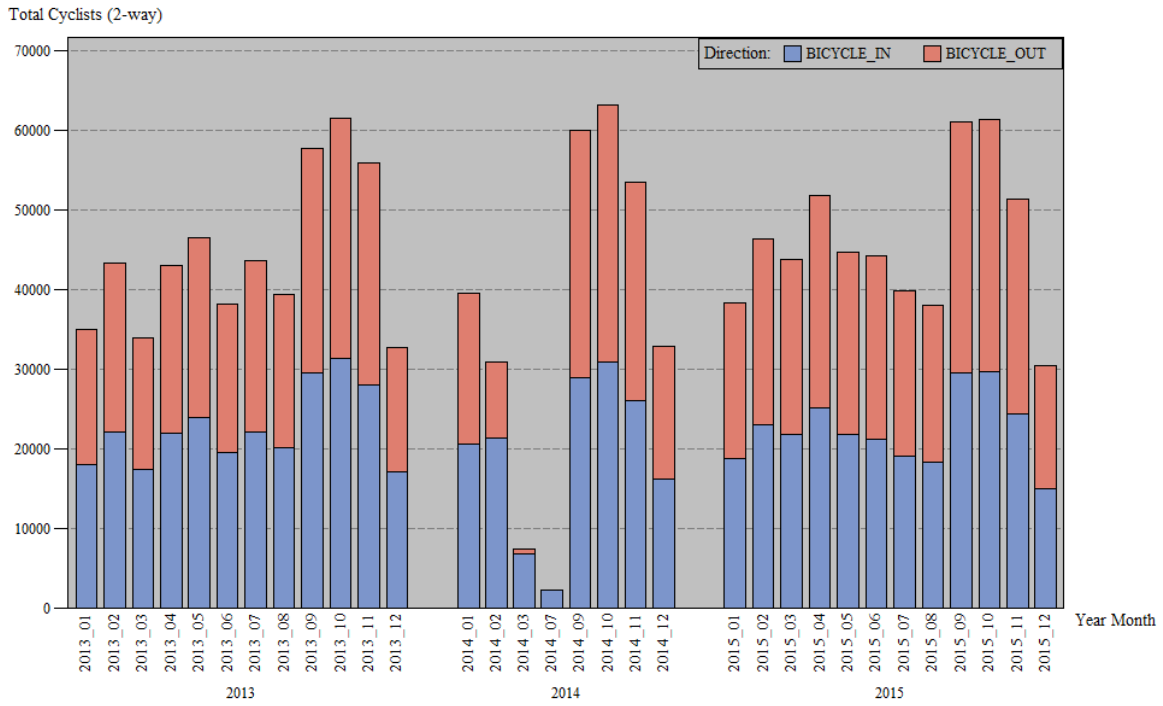
Location 5: Rock Road Inside park near the Dart (2-way)



	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	1,311	6,872	10,002	6,790	10,791	8,371	9,222	10,204	9,933	8,241	6,754	4,549
2013	8,006	8,248	5,443	8,648	11,031	13,571	20,959	15,575	15,632	12,583	11,732	8,164
2014	9,573	8,495	12,603	15,894	17,230	21,386	21,894	15,887	21,943	15,872	0	0
2015	0	0	0	16,983	16,201	21,129	17,517	18,027	18,649	16,393	12,487	8,614

Location 5 – "Rock Road Inside park near the Dart" is a two-way location, therefore a directional breakdown can be plotted. The data quality for this location is poor, with November 2014 – April 2015 completely missing. Although all the other months are populated the high variation between months potentially implies incorrect data – which can be just as damaging as missing values.

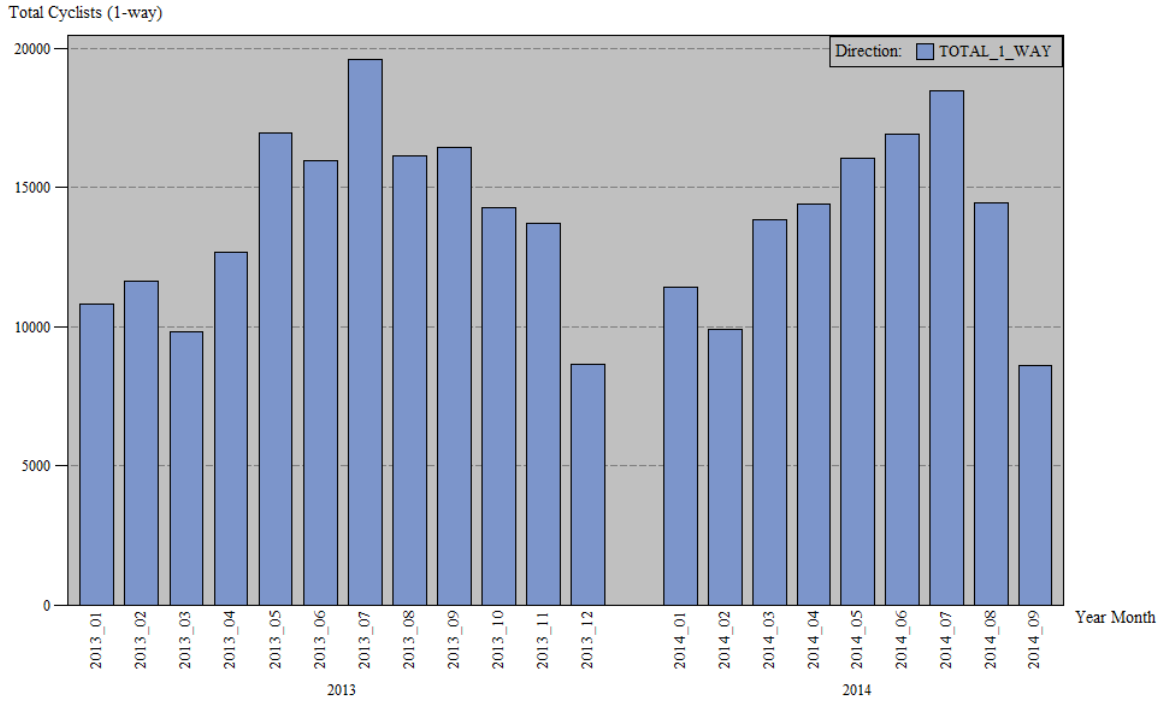
Location 6: Clonskeagh Road Totem (2-way)



	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	0	0	0	0	0	0	0	0	0	0	0	0
2013	34,937	43,268	33,937	42,944	46,475	38,172	43,567	39,364	57,647	61,514	55,857	32,686
2014	39,508	30,807	7,427	0	0	0	2,200	0	59,909	63,132	53,425	32,834
2015	38,257	46,346	43,673	51,810	44,633	44,222	39,773	37,896	61,077	61,259	51,271	30,336

Location 6 - Clonskeagh Road Totem is a two-way location, therefore a directional breakdown can be plotted. The data quality for this location is poor, with no data available for 2012 and several months missing for 2014, similar to previous locations high variation between months may imply incorrect data.

Location 7: Rock Road Totem (1-way)

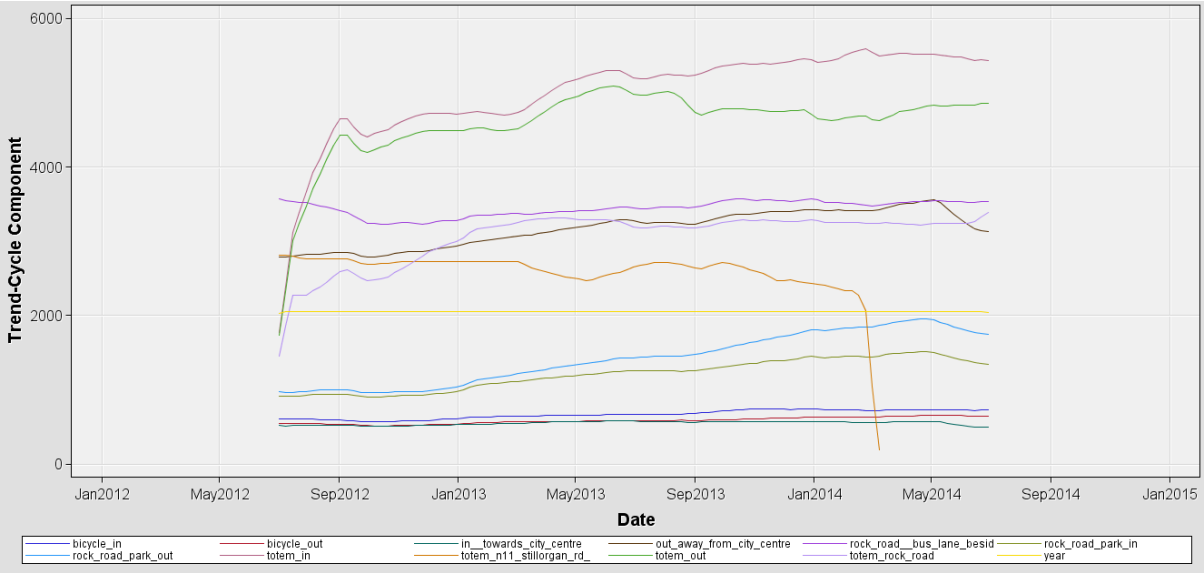


	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	0	0	0	0	0	0	0	0	0	0	0	0
2013	10,800	11,609	9,811	12,664	16,949	15,947	19,581	16,147	16,455	14,275	13,725	8,628
2014	11,415	9,878	13,833	14,388	16,050	16,926	18,461	14,428	8,581	0	0	0
2015	0	0	0	0	0	0	0	0	0	0	0	0

Location 7 - Rock Road Totem is a "one-way" location so directional data is not available. Similar to previous locations the data quality is very poor with no data available for 2012, 2015 and the last three months of 2014 missing.

c. Observations rationalised

Throughout all the plots, even those with exceedingly bad data quality a clear seasonal trend is observed. This is relatively straight forward to rationalise, people tend not to cycle during our cold and dark winters. The key positive trend observed is that there is a very slight upward trend in the total number of cyclists across most locations (see plot below), this is very welcome news considering we're facing into an obesity crisis. The elephant in the room is data quality, three of the seven locations have mission critical data quality issues, if this was a commercial project a go/no-go decision would need to be made at this point on whether to proceed or not with the modelling stage.



Part 2 - Predict the number of cyclists each month at each counter

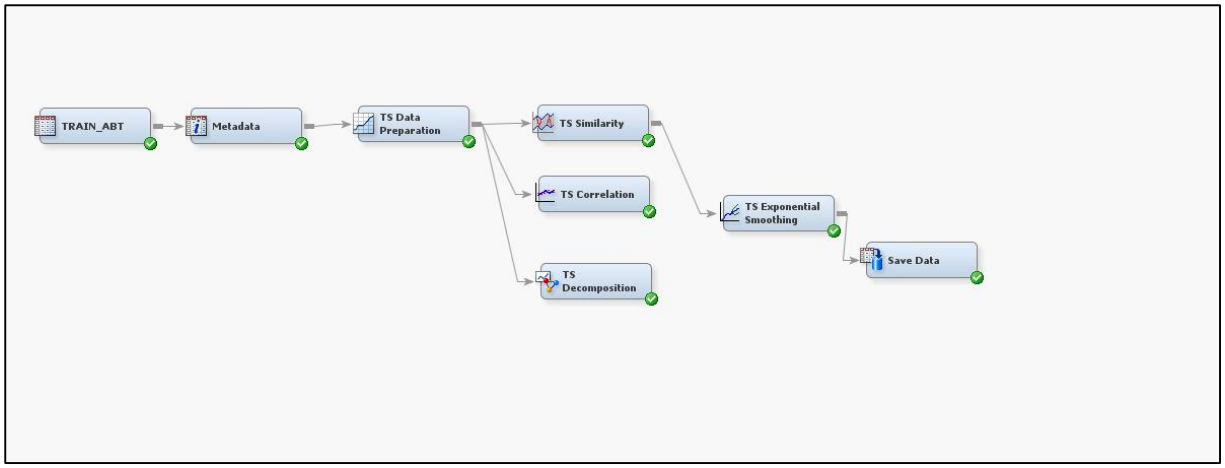
a. Methodology

Given the seasonal patterns observed following the data understanding step additional openly available data was sourced from met eireann including temperature, rainfall, sun hours and wind metrics. This data was at a granular hourly level, observation between 9pm and 6am were excluded before weekly averages were calculated. This data was merged with existing data to form a Modelling Analytics Base Table (ABT).

The forecast granularity (the level of aggregation) has a strong impact on the forecast accuracy. Take sales for example: it is easier to forecast the total sales for all products for a month across all geographies than to forecast on a weekly basis at a product level for individual geographies. As a rule of thumb, forecast accuracy decreases as granularity increases. The challenge with forecasting at a high level is that we begin to taper some trends and seasonal effects, therefore for this assignment it was decided to forecast at a weekly, directional (where available) level – aggregating up to a monthly location level after. This approach should help to reduce the variance due to missing data and incorrect data

Two models were considered for forecasting, the first ARMA with Regressors and the second Seasonal Exponential Smoothing (which is a special case of ARMA)

- ARIMA models are considered a very good general class of models that include a number of special cases. Typically an ARIMA model works well when the a time series can be made stationary (via differencing) or at least 4 full seasons of data is available
- Seasonal Exponential Smoothing - is an extension of exponential smoothing and therefore a special case of ARIMA. This approach works simultaneously estimates time-varying level, trend, and seasonal factors using recursive equations (therefore you do not need to adjust for seasonality).

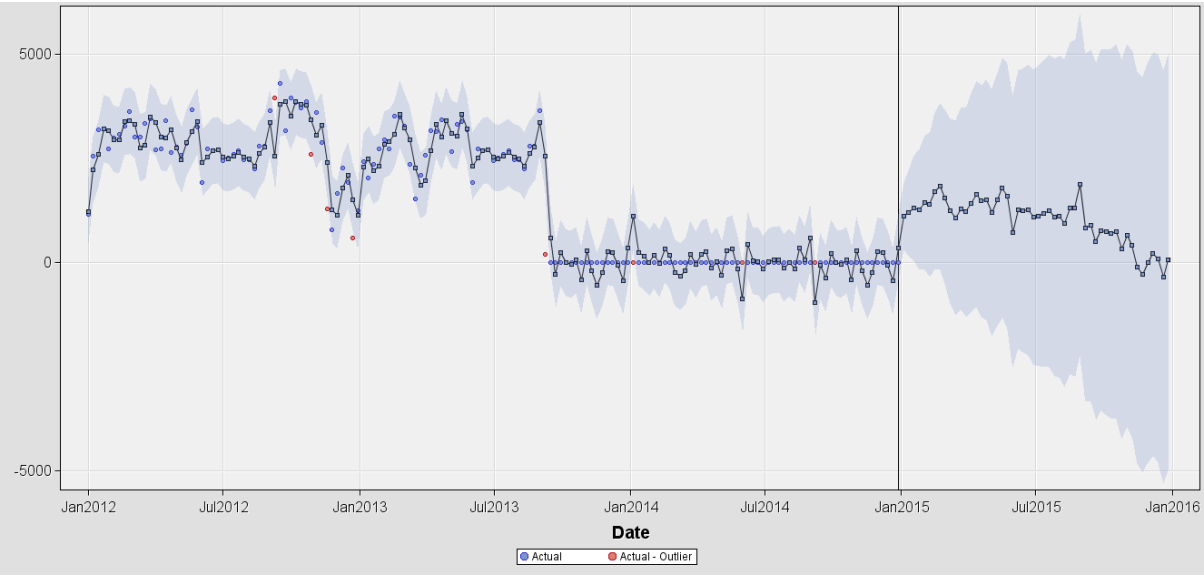


b. Results

Of the two models considered the Seasonal Exponential Smoothing proved to be most successful with the highest r-squared value and the lowest overall mean squared error. The performance of this model will be the primary focus of the discussion in the results section however some particularly interesting results from ARIMA model will also be discussed at the end of the section.

Location 1: N11 Stillorgan Road Totem (1-way)

The missing data for location 1 for both 2014 and 2015 make any form of prediction/evaluation near impossible. Potentially removing the nulled 2014 may have improved this forecast however it would be impossible to evaluation given the missing 2015 data.



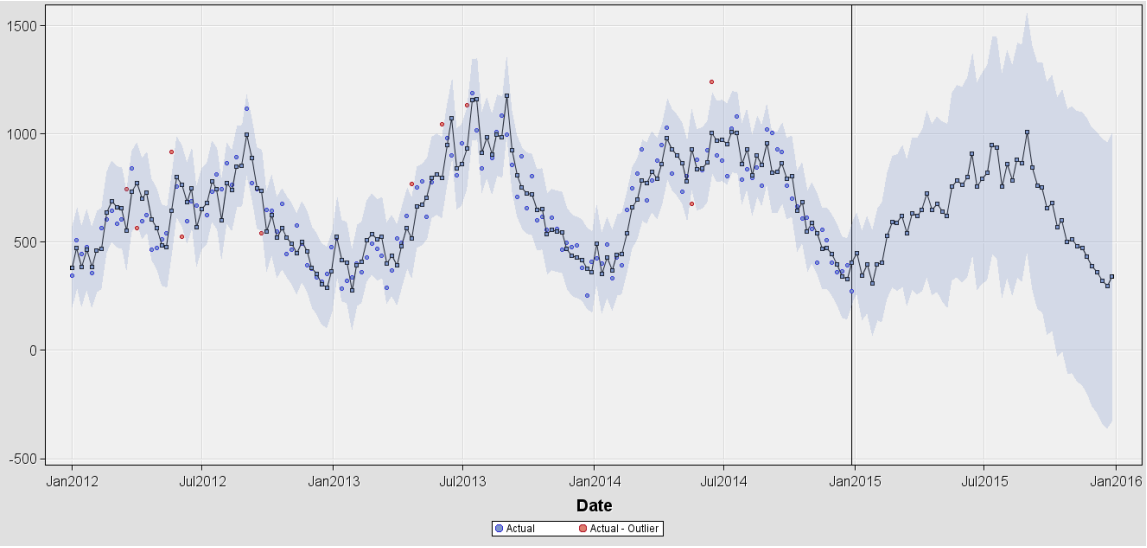
One-way

year_month	Actual	Prediction	Accuracy
2015_01	0	1767	0%
2015_02	0	3311	0%
2015_03	0	2653	0%
2015_04	0	3144	0%
2015_05	0	3302	0%
2015_06	0	2110	0%
2015_07	0	1939	0%
2015_08	0	3291	0%
2015_09	0	464	0%
2015_10	0	12	0%
2015_11	0	-2319	0%
2015_12	0	-2994	0%

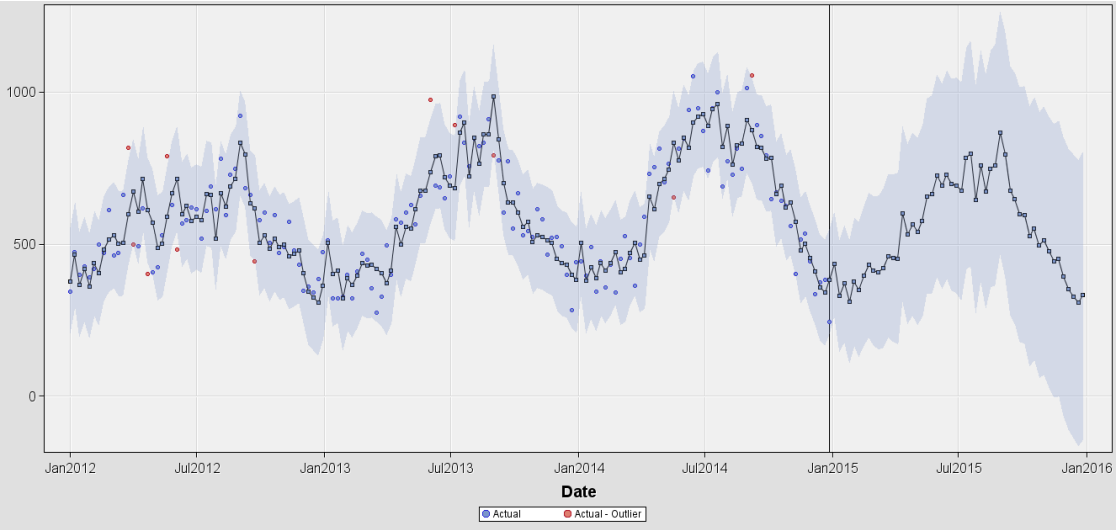
Location 2: The Metals, Glenageary Dart Station

The data quality for location 2 is probably the best of the seven locations resulting in a really strong forecast.

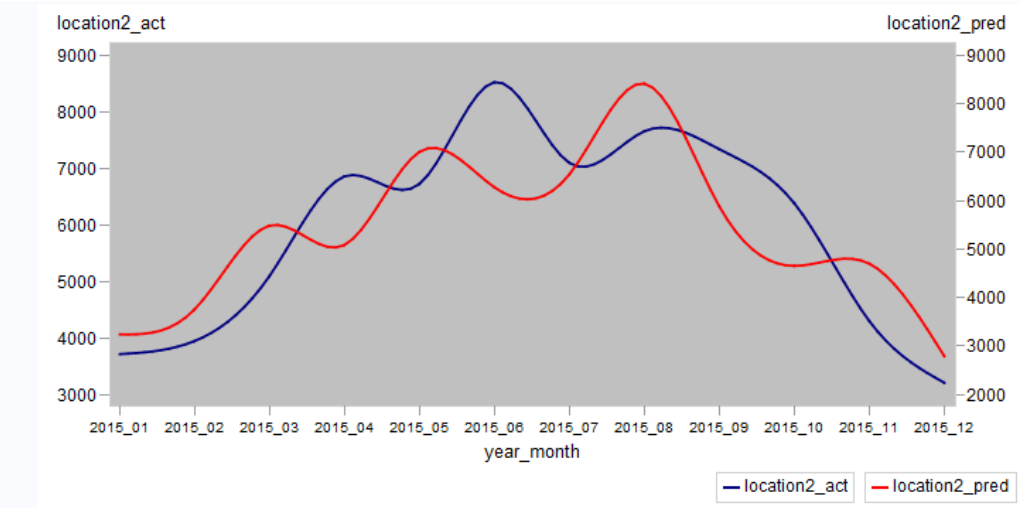
Direction	R-Square	Adjusted R-Square	Mean Square Error	Mean Absolute Error	Akaike Information Criterion
Inbound	0.818	0.817	9064.610	75.602	1434.605
Outbound	0.775	0.774	7704.530	68.595	1409.082



Inbound



Outbound



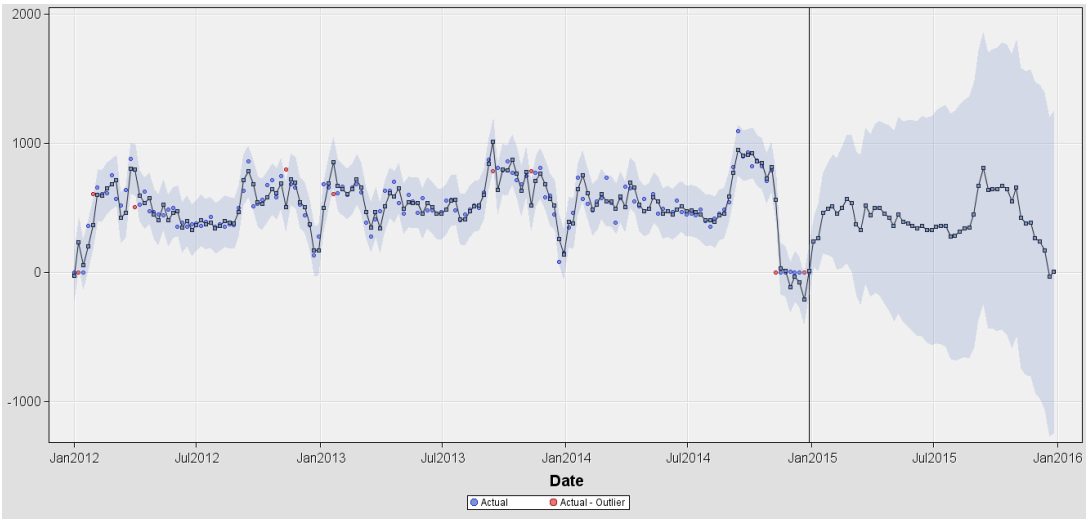
Actual vs Predicted

year_month	Actual	Prediction	Accuracy
2015_01	3,713	3,240	85%
2015_02	3,945	3,756	95%
2015_03	5,098	5,482	93%
2015_04	6,855	5,085	65%
2015_05	6,725	7,004	96%
2015_06	8,519	6,270	64%
2015_07	7,091	6,544	92%
2015_08	7,654	8,408	91%
2015_09	7,333	5,869	75%
2015_10	6,374	4,651	63%
2015_11	4,292	4,695	91%
2015_12	3,205	2,783	85%

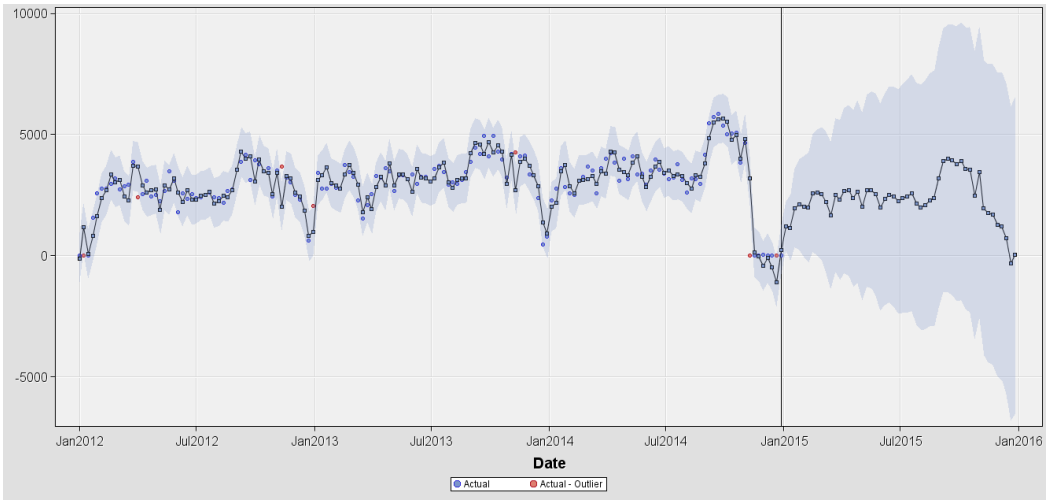
Location 3: N11 Montrose (2-way cycle track)

The data quality for this location is questionable, particularly for the inbound count which is dwarfed by the outbound count. The forecast looks relatively robust, the first 3 months of the out-of-time test set are missing so the evaluation is jeopardised.

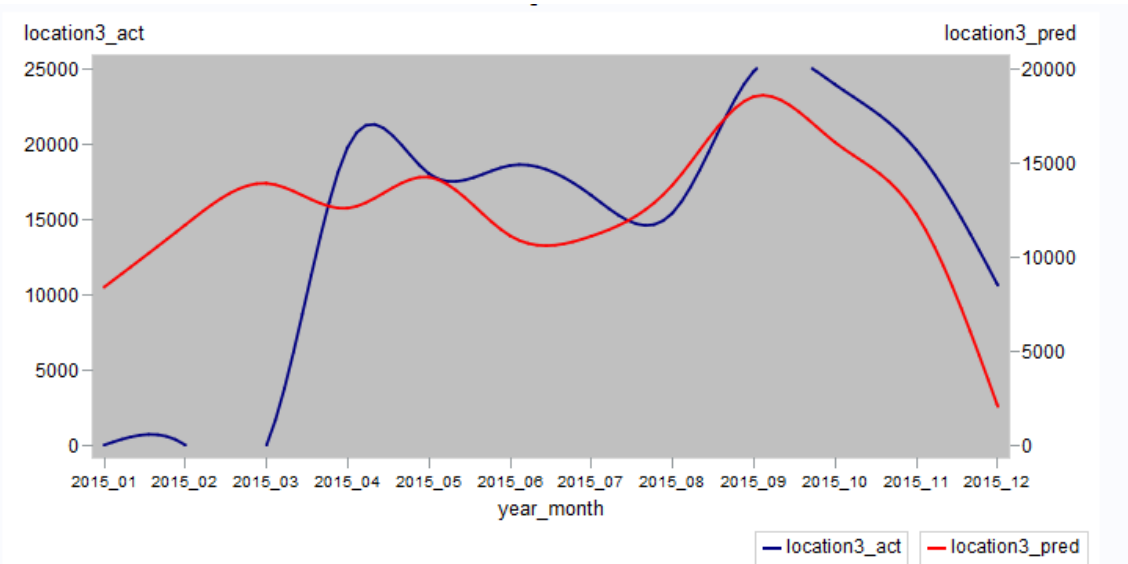
Direction	R-Square	Adjusted R-Square	Mean Square Error	Mean Absolute Error	Akaike Information Criterion
Inbound	0.780	0.779	9976.722	67.794	1449.658
Outbound	0.810	0.809	266797.887	351.195	1965.597



Inbound



Outbound



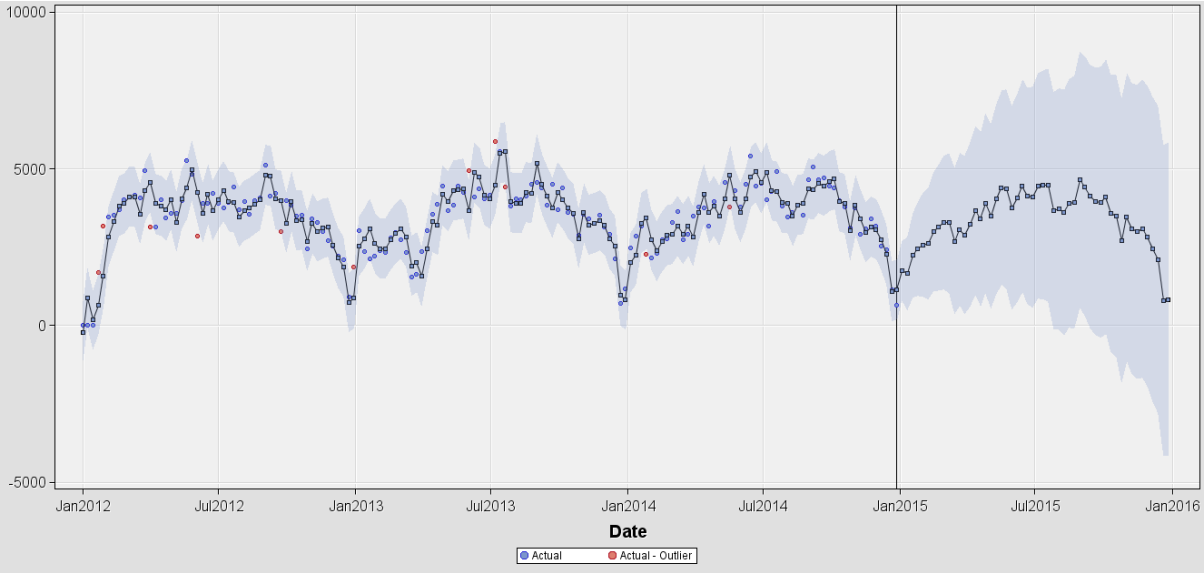
Actual vs Predicted

year_month	Actual	Prediction	Accuracy
2015_01	0	8390	0%
2015_02	0	11699	0%
2015_03	0	13905	0%
2015_04	19798	12585	43%
2015_05	18017	14235	73%
2015_06	18569	11113	33%
2015_07	16573	11102	51%
2015_08	15434	13838	88%
2015_09	24841	18516	66%
2015_10	23933	16073	51%
2015_11	19578	12245	40%
2015_12	10617	2055	-317%

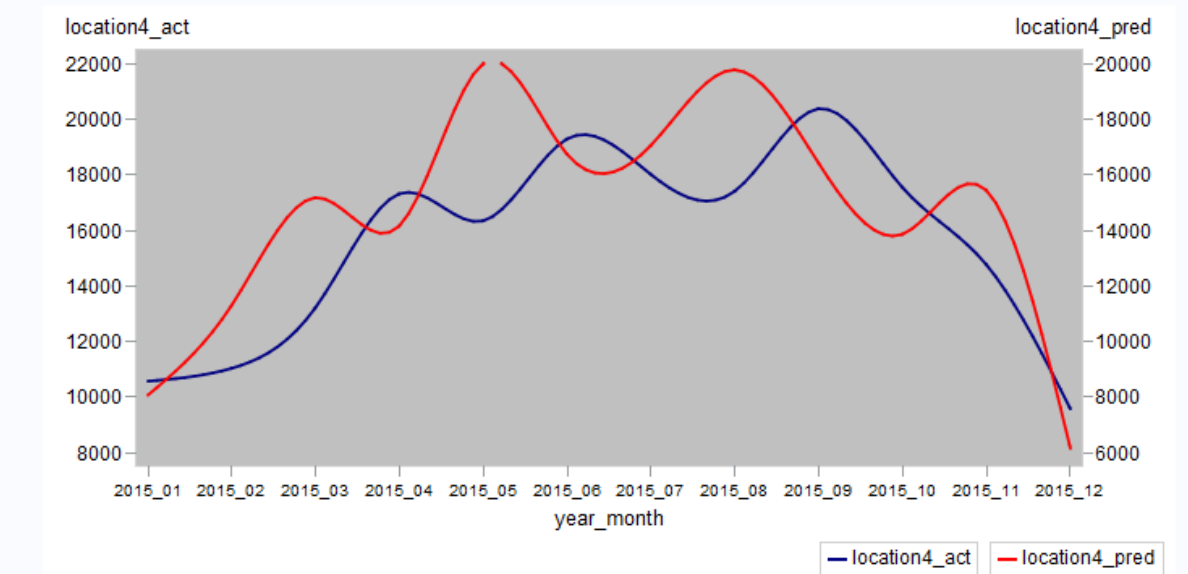
Location 4: Rock Road Bus lane (1-way)

With the exception of a number of outlier months in 2012 the data quality is relatively good with a clear seasonal trend being observed. The forecast again is relatively robust with a string r-squared value.

Direction	R-Square	Adjusted R-Square	Mean Square Error	Mean Absolute Error	Akaike Information Criterion
One-way	0.793	0.790	233251.905	355.779	1946.500



One -way



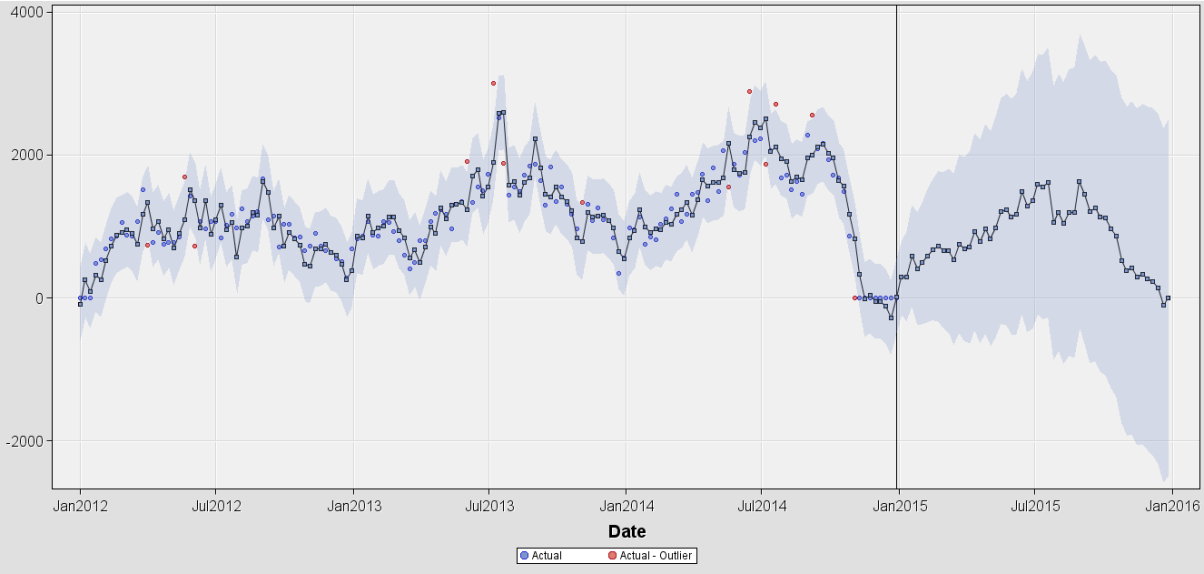
Actual vs Predicted

year_month	Actual	Prediction	Accuracy
2015_01	10571	8071	69%
2015_02	11037	11310	98%
2015_03	13229	15175	87%
2015_04	17317	14167	78%
2015_05	16344	19993	82%
2015_06	19292	16730	85%
2015_07	18005	17065	94%
2015_08	17415	19770	88%
2015_09	20375	16407	76%
2015_10	17531	13862	74%
2015_11	14756	15429	96%
2015_12	9574	6153	44%

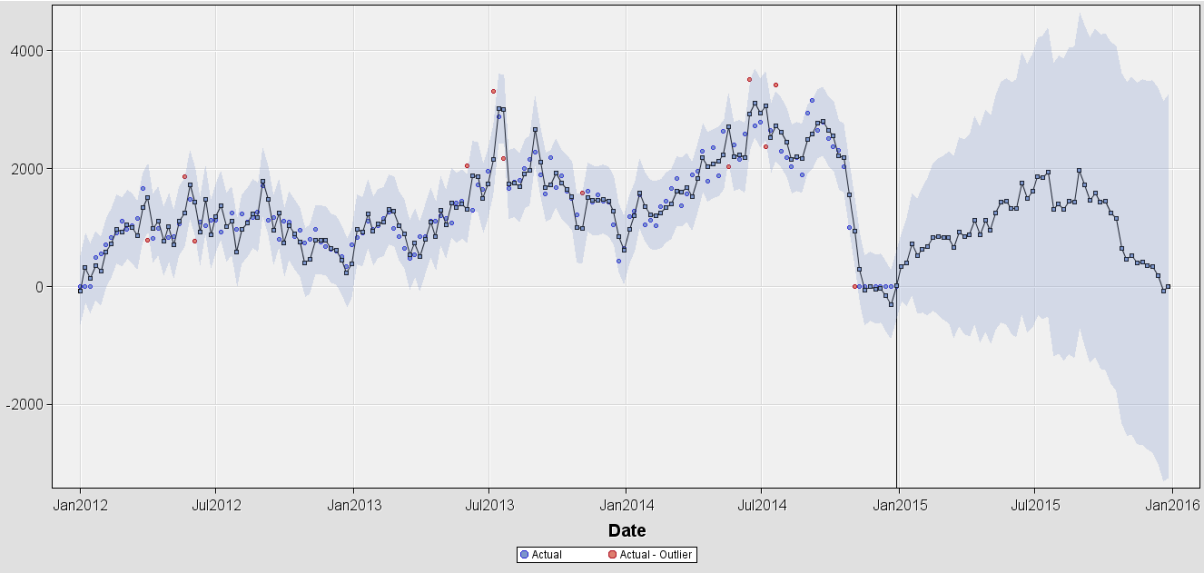
Location 5: Rock Road Inside park near the Dart (2-way)

The data quality for this location is poor, with November 2014 – April 2015 completely missing. Although all the other months are populated the high variation between months potentially implies incorrect data. Similar to previous locations data is missing from the test set, this jeopardises the forecast evaluation.

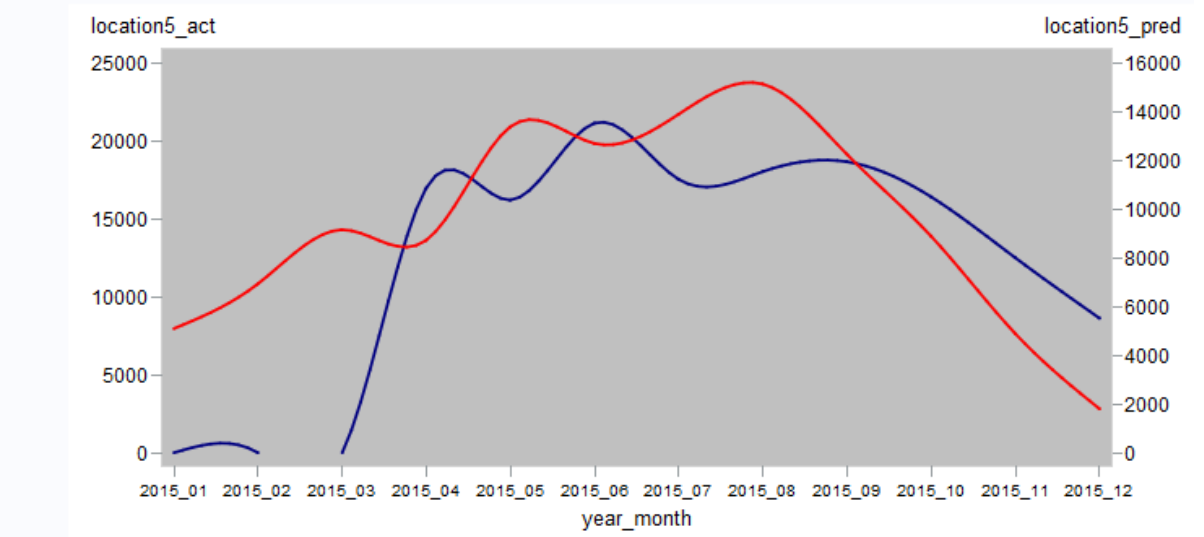
Direction	R-Square	Adjusted R-Square	Mean Square Error	Mean Absolute Error	Akaike Information Criterion
Inbound	0.730	0.779	9776.722	67.794	1449.658
Outbound	0.810	0.809	2697.887	351.195	1965.597



Inbound



Outbound



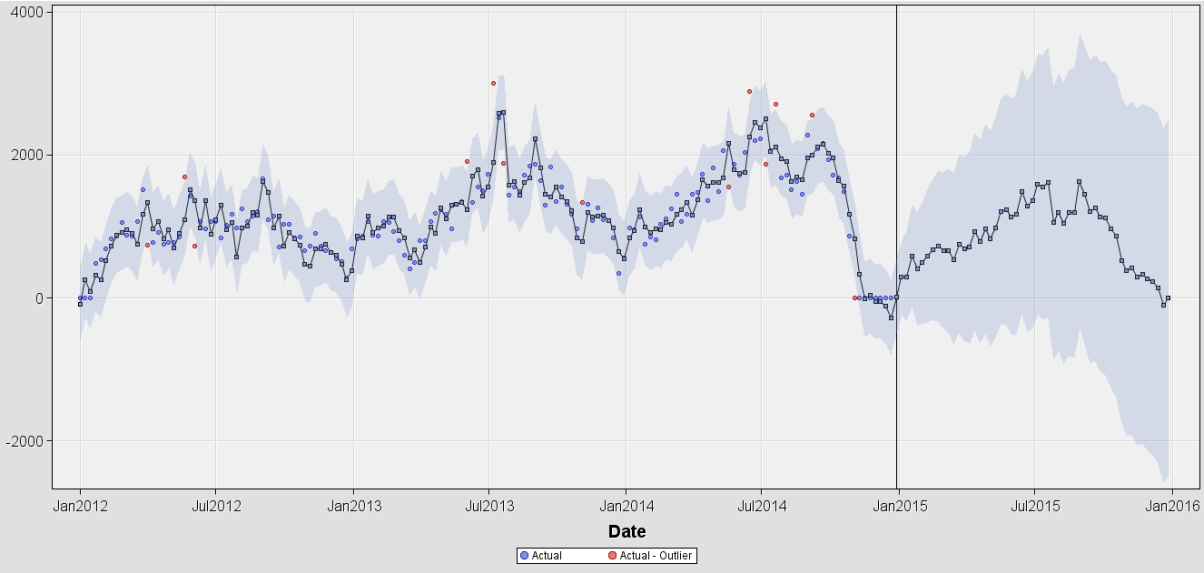
Actual vs Predicted

year_month	Actual	Prediction	Accuracy
2015_01	0	5094	0%
2015_02	0	6935	0%
2015_03	0	9149	0%
2015_04	16983	8724	5%
2015_05	16201	13379	79%
2015_06	21129	12687	33%
2015_07	17517	13893	74%
2015_08	18027	15120	81%
2015_09	18649	12229	48%
2015_10	16393	8870	15%
2015_11	12487	4877	-56%
2015_12	8614	1797	-279%

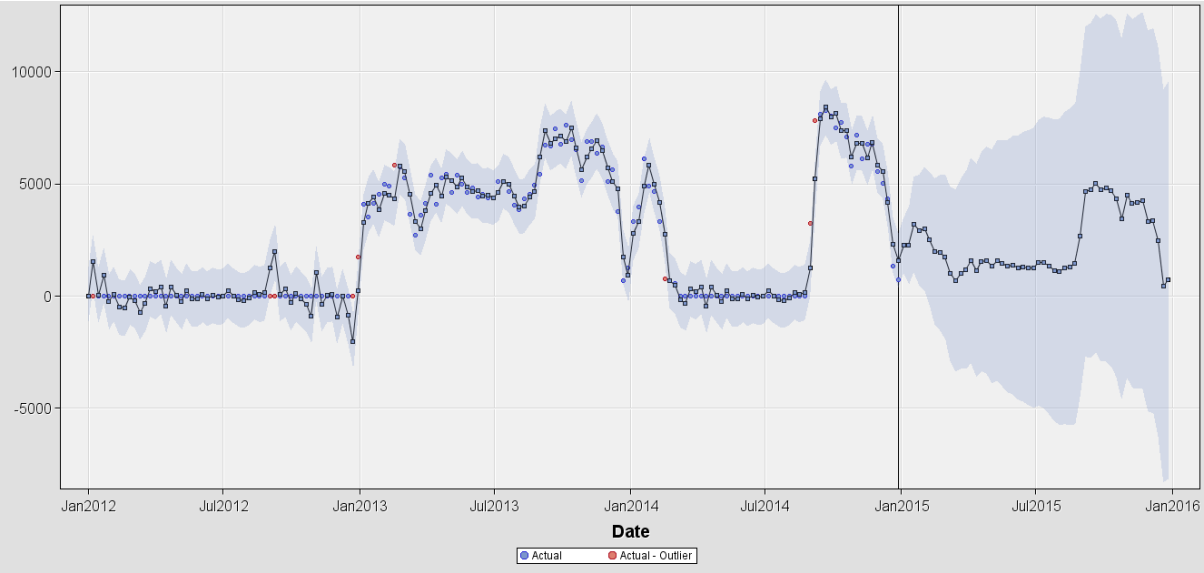
Location 6: Clonskeagh Road Totem (2-way)

The data quality for this location is poor, with no data available for 2012 and several months missing for 2014, similar to previous locations high variation between months may imply incorrect data.

Direction	R-Square	Adjusted R-Square	Mean Square Error	Mean Absolute Error	Akaike Information Criterion
Inbound	0.934	0.933	544201.600	483.257	2077.511
Outbound	0.951	0.951	387228.035	428.648	2024.083



Inbound

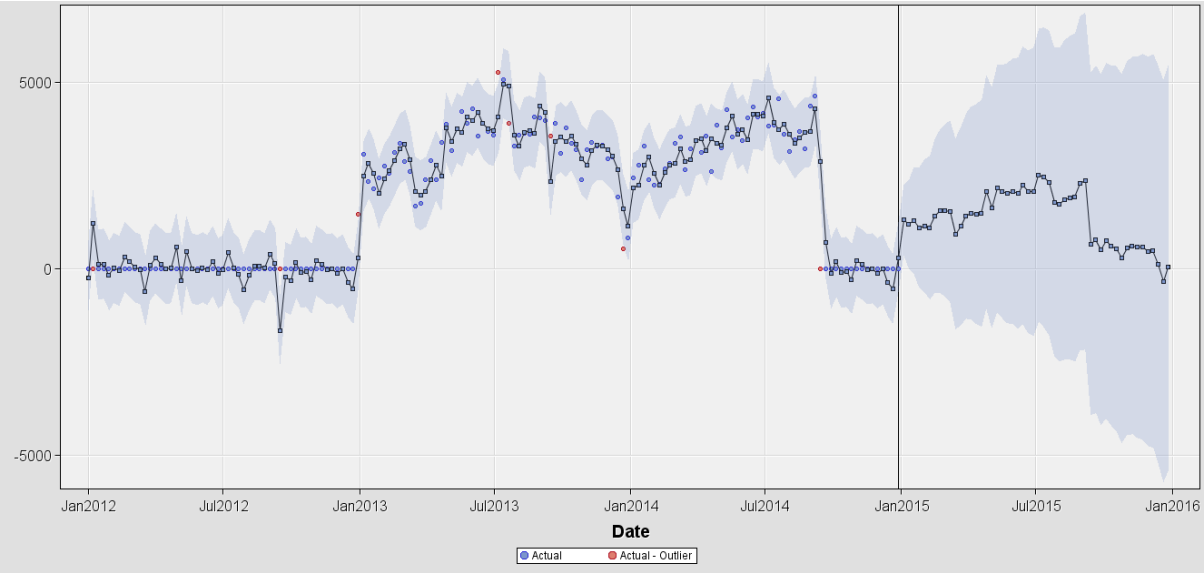


Outbound

year_month	Actual	Prediction	Accuracy
2015_01	0	8390	0%
2015_02	0	11699	0%
2015_03	0	13905	0%
2015_04	19798	12585	43%
2015_05	18017	14235	73%
2015_06	18569	11113	33%
2015_07	16573	11102	51%
2015_08	15434	13838	88%
2015_09	24841	18516	66%
2015_10	23933	16073	51%
2015_11	19578	12245	40%
2015_12	10617	2055	-317%

Location 7: Rock Road Totem (1-way)

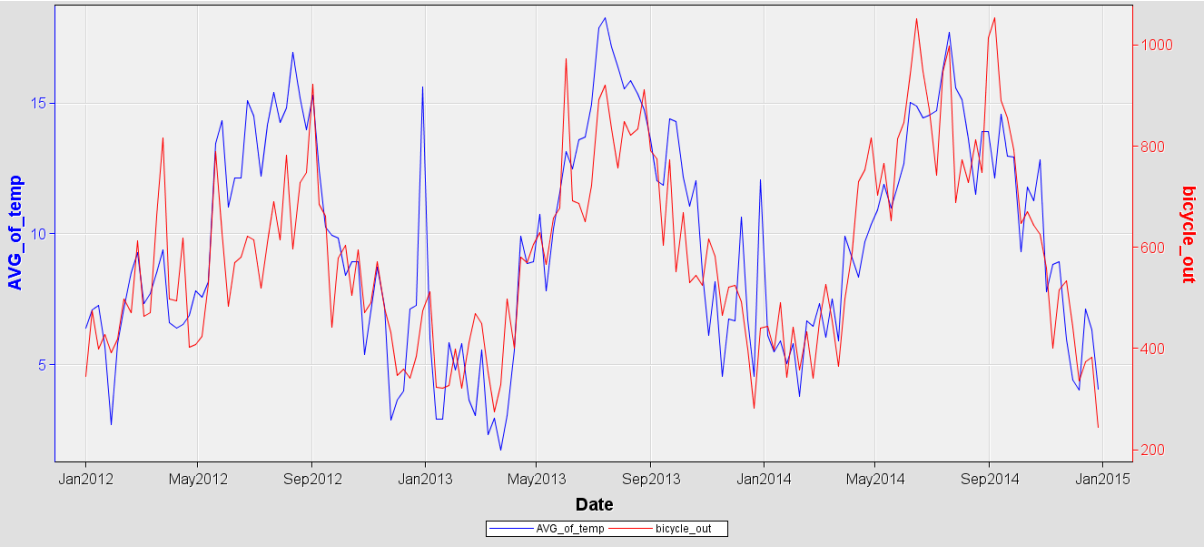
Similar to previous locations the data quality is very poor with no data available for 2012, 2015 and the last three months of 2014 missing.



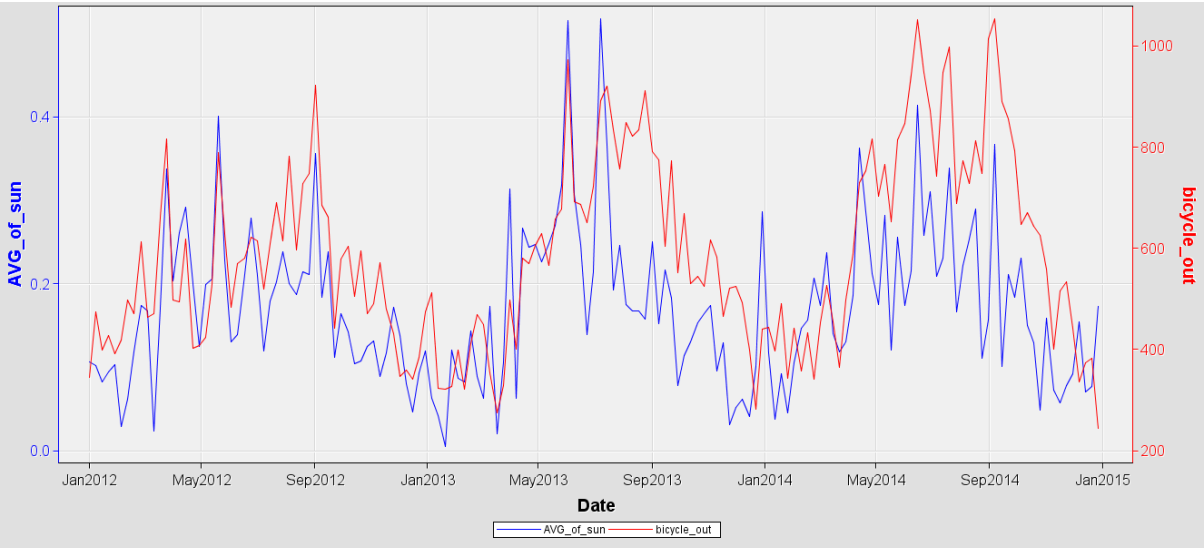
year_month	Actual	Prediction	Accuracy
2015_01	0	7627.263909	0%
2015_02	0	7838.370997	0%
2015_03	0	9731.730125	0%
2015_04	0	8915.062675	0%
2015_05	0	12814.42652	0%
2015_06	0	10591.41981	0%
2015_07	0	11158.24514	0%
2015_08	0	12173.65498	0%
2015_09	0	6203.935358	0%
2015_10	0	3954.761057	0%
2015_11	0	4891.886876	0%
2015_12	0	2471.706185	0%

ARIMA Regressors

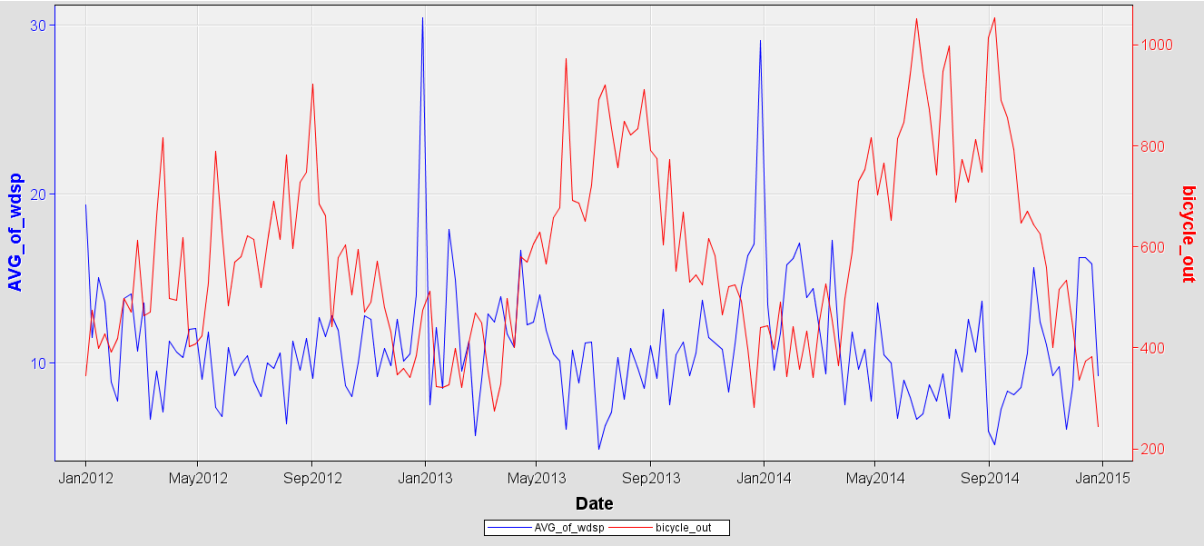
The following are a number of regressors considered for the ARIMA model, unfortunately none of these time series “lead” our target(s) enough to be significant predictors.



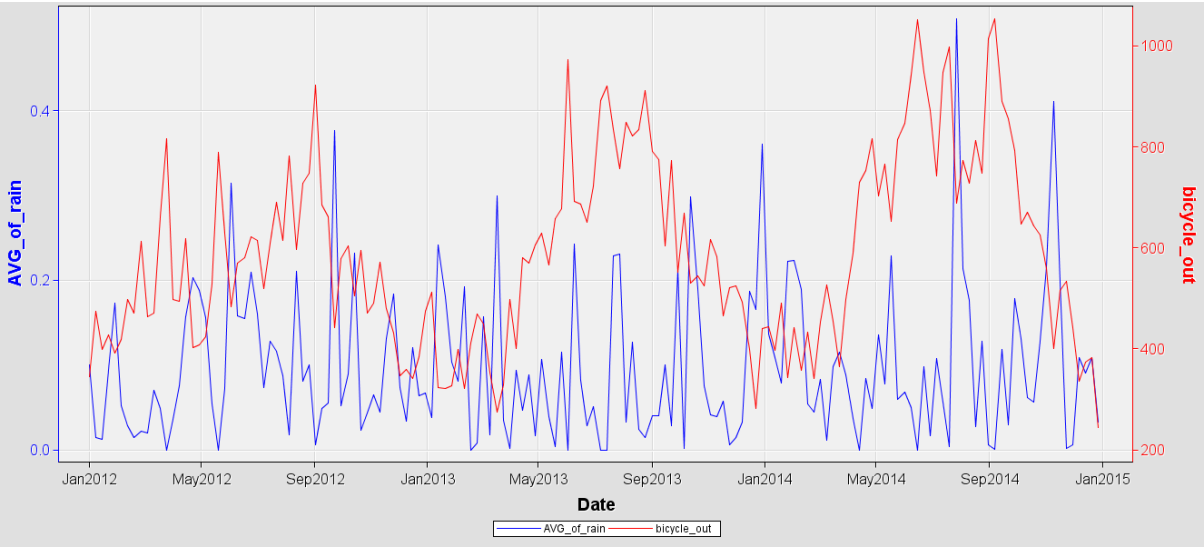
Average monthly temperature(C) vs counter



Average monthly sunshine (hours) temperature vs counter



Average monthly wind speed (knots) vs counter



Average monthly rainfall (mm) vs counter

c. Discussion

One of the key learning from this assignment is that a model is only as good as the data it learns from - garbage in garbage out. If this was a commercial project it is unlikely that models developed for locations 1,3 and 6 could be deployed as their robustness is questionable.

The assignment dataset was very limited therefore additional data was sourced from met eireann, however once the trends were plotted it became apparent that weather is quite unpredictable. The temperature and sunshine time series were rather strongly correlated with the counter time series but did not “lead” it in any significant way – therefore it was not a strong predictor.

The forecast granularity (the level of aggregation) has also being an interesting consideration for this assignment, you could imagine in the world of supply chain analytics selecting the correct level of granularity being a key challenge in order to maximise forecast accuracy and drive efficiencies.

Appendix

```
* PROJECT:           Assignment 3: Bicycle Counter Data
* NAME:              Read_in_data.sas
* AUTHOR:            Shane McCarthy
* EMAIL:             shane.mc-carthy@ucdconnect.ie
* DATE CREATED:      18/04/16
* PURPOSE:           This script reads in the data from csv files, ensures all
                     variables are of correct type and added the variable
*                    descriptions to the metadata as labels
;

libname wd "C:\Users\shane.mc.carthy\Dropbox\Masters\Semester2\MIS40970 Data Mining
for Bus Analytics\Assignments\Assignment 3\Data";

DATA wd.data_2012;
    LENGTH
        date               8
        glenageary         8
        bicycle_in         8
        bicycle_out        8
        walking_in         8
        walking_out        8
        n11_montrose       8
        out_away_from_city_centre   8
        in__towards_city_centre   8
        rock_road__bus_lane_beside_park   8
        rock_road__park    8
        rock_road_park_in   8
        rock_road_park_out   8
        totem_n11_stillorgan_rd_   8 ;
    LABEL
        date             = "Date"
        glenageary       = "DLR Co. Co. Glenageary"
        bicycle_in       = "Bicycle_IN"
        bicycle_out      = "Bicycle_OUT"
        walking_in       = "Walking_IN"
        walking_out      = "Walking_OUT"
        n11_montrose     = "N11 Montrose"
        out_away_from_city_centre = "OUT-Away From City Centre"
        in__towards_city_centre = "IN- Towards City Centre"
        rock_road__bus_lane_beside_park = "Rock Road *Bus Lane Beside Park"
        rock_road__park  = "Rock Road *Park"
        rock_road_park_in = "Rock_Road_Park_IN"
        rock_road_park_out = "Rock_Road_Park_OUT"
        totem_n11_stillorgan_rd_ = "Totem N11 Stillorgan Rd." ;
    FORMAT
        date             DATE9.
        glenageary       BEST12.
        bicycle_in       BEST12.
        bicycle_out      BEST12.
        walking_in       BEST12.
        walking_out      BEST12.
        n11_montrose     BEST12.
        out_away_from_city_centre BEST12.
        in__towards_city_centre BEST12.
        rock_road__bus_lane_beside_park BEST12.
        rock_road__park  BEST12.
        rock_road_park_in BEST12.
        rock_road_park_out BEST12.
        totem_n11_stillorgan_rd_ BEST12. ;
    INFORMAT
        date             DATE9.
```

```
        glenageary       BEST12.
        bicycle_in       BEST12.
        bicycle_out      BEST12.
        walking_in       BEST12.
        walking_out      BEST12.
        n11_montrose     BEST12.
        out_away_from_city_centre BEST12.
        in__towards_city_centre BEST12.
        rock_road__bus_lane_beside_park BEST12.
        rock_road__park  BEST12.
        rock_road_park_in BEST12.
        rock_road_park_out BEST12.
        totem_n11_stillorgan_rd_ BEST12. ;
    INFILE 'C:\Users\shane.mc.carthy\AppData\Local\Temp\SEG868\data-
b99f7338a2c54d1097da3d2e8c8f4dc0.txt'
        LRECL=57
        ENCODING="WLATIN1"
        TERMSTR=CRLF
        DLM='7F'x
        MISSOVER
        DSD ;
    INPUT
        date             : BEST32.
        glenageary       : BEST32.
        bicycle_in       : BEST32.
        bicycle_out      : BEST32.
        walking_in       : BEST32.
        walking_out      : BEST32.
        n11_montrose     : BEST3.
        out_away_from_city_centre : BEST3.
        in__towards_city_centre : BEST3.
        rock_road__bus_lane_beside_park : BEST3.
        rock_road__park  : BEST3.
        rock_road_park_in : BEST3.
        rock_road_park_out : BEST3.
        totem_n11_stillorgan_rd_ : BEST32. ;
RUN;



DATA wd.data_2013;
    LENGTH
        date               8
        glenageary         8
        bicycle_in         8
        bicycle_out        8
        walking_in         8
        walking_out        8
        n11_montrose       8
        out_away_from_city_centre   8
        in__towards_city_centre   8
        rock_road__bus_lane_beside_park   8
        rock_road__park    8
        rock_road_park_in   8
        rock_road_park_out   8
        totem_clonskeagh_road   8
        totem_in           8
        totem_out          8
        totem_n11_stillorgan_rd_   8
        totem_rock_road    8 ;
    LABEL
        date             = "Date"
        glenageary       = "DLR Co. Co. Glenageary"
        bicycle_in       = "Bicycle_IN"
        bicycle_out      = "Bicycle_OUT"
        walking_in       = "Walking_IN"
```

```
        walking_out      = "Walking_OUT"
        n11_montrose     = "N11 Montrose"
        out_away_from_city_centre = "OUT-Away From City Centre"
        in__towards_city_centre = "IN- Towards City Centre"
        rock_road__bus_lane_beside_park = "Rock Road *Bus Lane Beside Park"
        rock_road__park  = "Rock Road *Park"
        rock_road_park_in = "Rock_Road_Park_IN"
        rock_road_park_out = "Rock_Road_Park_OUT"
        totem_clonskeagh_road = "Totem Clonskeagh Road"
        totem_in         = "Totem IN"
        totem_out        = "Totem OUT"
        totem_n11_stillorgan_rd_ = "Totem N11 Stillorgan Rd."
        totem_rock_road  = "Totem Rock Road" ;
    FORMAT
        date             DATE9.
        glenageary       BEST12.
        bicycle_in       BEST12.
        bicycle_out      BEST12.
        walking_in       BEST12.
        walking_out      BEST12.
        n11_montrose     BEST12.
        out_away_from_city_centre BEST12.
        in__towards_city_centre BEST12.
        rock_road__bus_lane_beside_park BEST12.
        rock_road__park  BEST12.
        rock_road_park_in BEST12.
        rock_road_park_out BEST12.
        totem_clonskeagh_road BEST12.
        totem_in         BEST12.
        totem_out        BEST12.
        totem_n11_stillorgan_rd_ BEST12.
        totem_rock_road  BEST12. ;
    INFORMAT
        date             DATE9.
        glenageary       BEST12.
        bicycle_in       BEST12.
        bicycle_out      BEST12.
        walking_in       BEST12.
        walking_out      BEST12.
        n11_montrose     BEST12.
        out_away_from_city_centre BEST12.
        in__towards_city_centre BEST12.
        rock_road__bus_lane_beside_park BEST12.
        rock_road__park  BEST12.
        rock_road_park_in BEST12.
        rock_road_park_out BEST12.
        totem_clonskeagh_road BEST12.
        totem_in         BEST12.
        totem_out        BEST12.
        totem_n11_stillorgan_rd_ BEST12.
        totem_rock_road  BEST12. ;
    INFILE 'C:\Users\shane.mc.carthy\AppData\Local\Temp\SEG868\data-
a2ae05e0bcb54a3c8727cc8c3659b90d.txt'
        LRECL=77
        ENCODING="WLATIN1"
        TERMSTR=CRLF
        DLM='7F'x
        MISSOVER
        DSD ;
    INPUT
        date             : BEST32.
        glenageary       : BEST32.
        bicycle_in       : BEST32.
        bicycle_out      : BEST32.
        walking_in       : BEST32.
        walking_out      : BEST32.
        n11_montrose     : BEST32.
        out_away_from_city_centre : BEST32.
```

```
        in__towards_city_centre : BEST32.
        rock_road__bus_lane_beside_park : BEST32.
        rock_road__park  : BEST32.
        rock_road_park_in : BEST32.
        rock_road_park_out : BEST32.
        totem_clonskeagh_road : BEST4.
        totem_in         : BEST4.
        totem_out        : BEST4.
        totem_n11_stillorgan_rd_ : BEST3.
        totem_rock_road  : BEST4. ;
RUN;


DATA wd.data_2014;
    LENGTH
        date               8
        glenageary         8
        bicycle_in         8
        bicycle_out        8
        walking_in         8
        walking_out        8
        n11_montrose       8
        out_away_from_city_centre   8
        in__towards_city_centre   8
        rock_road__bus_lane_beside_park   8
        rock_road__park    8
        rock_road_park_in   8
        rock_road_park_out   8
        totem_clonskeagh_road   8
        totem_in           8
        totem_out          8
        totem_rock_road    8 ;
    LABEL
        date             = "Date"
        glenageary       = "DLR Co. Co. Glenageary"
        bicycle_in       = "Bicycle_IN"
        bicycle_out      = "Bicycle_OUT"
        walking_in       = "Walking_IN"
        walking_out      = "Walking_OUT"
        n11_montrose     = "N11 Montrose"
        out_away_from_city_centre = "OUT-Away From City Centre"
        in__towards_city_centre = "IN- Towards City Centre"
        rock_road__bus_lane_beside_park = "Rock Road *Bus Lane Beside Park"
        rock_road__park  = "Rock Road *Park"
        rock_road_park_in = "Rock_Road_Park_IN"
        rock_road_park_out = "Rock_Road_Park_OUT"
        totem_clonskeagh_road = "Totem Clonskeagh Road"
        totem_in         = "Totem IN"
        totem_out        = "Totem OUT"
        totem_rock_road  = "Totem Rock Road" ;
    FORMAT
        date             DATE9.
        glenageary       BEST12.
        bicycle_in       BEST12.
        bicycle_out      BEST12.
        walking_in       BEST12.
        walking_out      BEST12.
        n11_montrose     BEST12.
        out_away_from_city_centre BEST12.
        in__towards_city_centre BEST12.
        rock_road__bus_lane_beside_park BEST12.
        rock_road__park  BEST12.
        rock_road_park_in BEST12.
        rock_road_park_out BEST12.
        totem_clonskeagh_road BEST12.
        totem_in         BEST12.
        totem_out        BEST12.
        totem_rock_road  BEST12. ;
```

```
    INFORMAT
        date              DATE9.
        glenageary        BEST12.
        bicycle_in        BEST12.
        bicycle_out       BEST12.
        walking_in        BEST12.
        walking_out       BEST12.
        n11_montrose      BEST12.
        out_away_from_city_centre BEST12.
        in__towards_city_centre BEST12.
        rock_road__bus_lane_beside_park BEST12.
        rock_road__park  BEST12.
        rock_road_park_in BEST12.
        rock_road_park_out BEST12.
        totem_clonskeagh_road BEST12.
        totem_in          BEST12.
        totem_out         BEST12.
        totem_rock_road  BEST12. ;
    INFILE 'C:\Users\shane.mc.carthy\AppData\Local\Temp\SEG868\data-
17686c09f76c4334a538872a58612f90.txt'
        LRECL=74
        ENCODING="WLATIN1"
        TERMSTR=CRLF
        DLM='7F'x
        MISSOVER
        DSD ;
    INPUT
        date              : BEST32.
        glenageary        : BEST32.
        bicycle_in        : BEST32.
        bicycle_out       : BEST32.
        walking_in        : BEST32.
        walking_out       : BEST32.
        n11_montrose      : BEST32.
        out_away_from_city_centre : BEST32.
        in__towards_city_centre : BEST32.
        rock_road__bus_lane_beside_park : BEST32.
        rock_road__park  : BEST32.
        rock_road_park_in : BEST32.
        rock_road_park_out : BEST32.
        totem_clonskeagh_road : BEST4.
        totem_in          : BEST4.
        totem_out         : BEST4.
        totem_rock_road  : BEST3. ;
RUN;



DATA wd.data_2015;
    LENGTH
        date               8
        glenageary         8
        walking_in         8
        walking_out        8
        bicycle_in         8
        bicycle_out        8
        n11_eco_totem      8
        n11_montrose       8
        out_away_from_city_centre   8
        in__towards_city_centre   8
        rock_road__bus_lane_beside_park   8
        rock_road__park    8
        rock_road_park_in   8
        rock_road_park_out   8
        rock_road_inbound   8
        totem_clonskeagh_road__data_only   8
        totem_in           8
```

```
        totem_out          8
        totem_n11_stillorgan_rd__data_on   8
        totem_rock_road__data_only_    8 ;
    LABEL
        date              = "Date"
        glenageary        = "Glenageary"
        walking_in        = "Walking_IN"
        walking_out       = "Walking_OUT"
        bicycle_in        = "Bicycle_IN"
        bicycle_out       = "Bicycle_OUT"
        n11_eco_totem     = "N11 ECO TOTEM"
        n11_montrose      = "N11 Montrose"
        out_away_from_city_centre = "OUT-Away From City Centre"
        in__towards_city_centre = "IN- Towards City Centre"
        rock_road__bus_lane_beside_park = "Rock Road *Bus Lane Beside Park"
        rock_road__park  = "Rock Road *Park"
        rock_road_park_in = "Rock_Road_Park_IN"
        rock_road_park_out = "Rock_Road_Park_OUT"
        rock_road_inbound = "Rock Road Inbound"
        totem_clonskeagh_road__data_only = "Totem Clonskeagh Road (Data Only)"
        totem_in          = "Totem IN"
        totem_out         = "Totem OUT"
        totem_n11_stillorgan_rd__data_on = "Totem N11 Stillorgan Rd (Data Only)"
        totem_rock_road__data_only_ = "Totem Rock Road (Data Only)" ;
    FORMAT
        date              DATE9.
        glenageary        BEST12.
        walking_in        BEST12.
        walking_out       BEST12.
        bicycle_in        BEST12.
        bicycle_out       BEST12.
        n11_eco_totem     BEST12.
        n11_montrose      BEST12.
        out_away_from_city_centre BEST12.
        in__towards_city_centre BEST12.
        rock_road__bus_lane_beside_park BEST12.
        rock_road__park  BEST12.
        rock_road_park_in BEST12.
        rock_road_park_out BEST12.
        rock_road_inbound BEST12.
        totem_clonskeagh_road__data_only BEST12.
        totem_in          BEST12.
        totem_out         BEST12.
        totem_n11_stillorgan_rd__data_on BEST12.
        totem_rock_road__data_only_ BEST12. ;
    INFORMAT
        date              DATE9.
        glenageary        BEST12.
        walking_in        BEST12.
        walking_out       BEST12.
        bicycle_in        BEST12.
        bicycle_out       BEST12.
        n11_eco_totem     BEST12.
        n11_montrose      BEST12.
        out_away_from_city_centre BEST12.
        in__towards_city_centre BEST12.
        rock_road__bus_lane_beside_park BEST12.
        rock_road__park  BEST12.
        rock_road_park_in BEST12.
        rock_road_park_out BEST12.
        rock_road_inbound BEST12.
        totem_clonskeagh_road__data_only BEST12.
        totem_in          BEST12.
        totem_out         BEST12.
        totem_n11_stillorgan_rd__data_on BEST12.
        totem_rock_road__data_only_ BEST12. ;
    INFILE 'C:\Users\shane.mc.carthy\AppData\Local\Temp\SEG868\data-
e4de0006195740bf8ed712637cefb399.txt'
```

```
        LRECL=78
        ENCODING="WLATIN1"
        TERMSTR=CRLF
        DLM='7F'x
        MISSOVER
        DSD ;
    INPUT
        date              : BEST32.
        glenageary        : BEST32.
        walking_in        : BEST32.
        walking_out       : BEST32.
        bicycle_in        : BEST32.
        bicycle_out       : BEST32.
        n11_eco_totem     : BEST3.
        n11_montrose      : BEST32.
        out_away_from_city_centre : BEST32.
        in__towards_city_centre : BEST32.
        rock_road__bus_lane_beside_park : BEST32.
        rock_road__park   : BEST32.
        rock_road_park_in : BEST32.
        rock_road_park_out : BEST32.
        rock_road_inbound : BEST4.
        totem_clonskeagh_road__data_only : BEST32.
        totem_in          : BEST32.
        totem_out         : BEST32.
        totem_n11_stillorgan_rd__data_on : BEST1.
        totem_rock_road__data_only_ : BEST1. ;
RUN;

%macro inspect_meta(num_sheet=,);

%DO n=1 %TO &num_sheet.;
%Put *** Reading in &n. of &num_file. sheets;

proc sql;
       create table part_%eval(2011+&n.) as
              select
                     upcase(name) as variable_%eval(2011+&n.)
              from dictionary.columns
                     where libname ="WD"
                            AND memname="DATA_20%eval(11+&n.)";
quit;
%end;


       data unique_variables;
       set part_2012(rename=(variable_2012=variable))
              part_2013(rename=(variable_2013=variable))
              part_2014(rename=(variable_2014=variable))
              part_2015(rename=(variable_2015=variable)) ;
       run;

       proc sort data=unique_variables nodup; by variable; run;

       data unique_variables;
       set unique_variables;
       ID =_N_;
       run;


              PROC SQL;
              CREATE TABLE compare_tables AS
              SELECT t1.ID,
              t1.variable
              ,case when t12.variable_2012 ="" then 0 else 1 end  as variable_2012
              ,case when t13.variable_2013 ="" then 0 else 1 end  as variable_2013
              ,case when t14.variable_2014 ="" then 0 else 1 end  as variable_2014
              ,case when t15.variable_2015 ="" then 0 else 1 end  as variable_2015
```

```
            ,case when calculated  variable_2012 + calculated variable_2013
                        + calculated variable_2014 +  calculated variable_2015 <
4 then 1 else 0 end as Error
            FROM WORK.UNIQUE_VARIABLES t1
            left join part_2012 t12 on t12.variable_2012=t1.variable
            left join part_2013 t13 on t13.variable_2013=t1.variable
            left join part_2014 t14 on t14.variable_2014=t1.variable
            left join part_2015 t15 on t15.variable_2015=t1.variable
             ;
QUIT;

%mend;
%inspect_meta(num_sheet=4);



data wd.all_counter_data;
set wd.data_2012 wd.data_2013 wd.data_2014 wd.data_2015;
run;




DATA wd.data_met;
    LENGTH
        Date               8
        irain              8
        rain               8
        itemp              8
        temp               8
        iwb                8
        wetb               8
        dewpt              8
        vappr              8
        rhum               8
        msl                8
        iwdsp              8
        wdsp               8
        iwddir             8
        wddir              8
        ww                 8
        w                  8
        sun                8
        vis                8
        clht               8
        clamt              8 ;
    FORMAT
        Date             DATETIME18.
        irain            BEST1.
        rain             BEST4.
        itemp            BEST1.
        temp             BEST5.
        iwb              BEST1.
        wetb             BEST5.
        dewpt            BEST5.
        vappr            BEST4.
        rhum             BEST3.
        msl              BEST6.
        iwdsp            BEST1.
        wdsp             BEST2.
        iwddir           BEST1.
        wddir            BEST3.
        ww               BEST2.
        w                BEST2.
        sun              BEST3.
```

```
        vis             BEST5.
        clht            BEST3.
        clamt           BEST1. ;
    INFORMAT
        Date            DATETIME18.
        irain           BEST1.
        rain            BEST4.
        itemp           BEST1.
        temp            BEST5.
        iwb             BEST1.
        wetb            BEST5.
        dewpt           BEST5.
        vappr           BEST4.
        rhum            BEST3.
        msl             BEST6.
        iwdsp           BEST1.
        wdsp            BEST2.
        iwddir          BEST1.
        wddir           BEST3.
        ww              BEST2.
        w               BEST2.
        sun             BEST3.
        vis             BEST5.
        clht            BEST3.
        clamt           BEST1. ;
    INFILE 'C:\Users\shane.mc.carthy\AppData\Local\Temp\SEG1496\hly532-
0511228317684745aa8c118a5fc31296.txt'
        LRECL=89
        ENCODING="WLATIN1"
        TERMSTR=CRLF
        DLM='7F'x
        MISSOVER
        DSD ;
    INPUT
        Date            : ?? ANYDTDTM16.
        irain           : ?? BEST1.
        rain            : ?? COMMA4.
        itemp           : ?? BEST1.
        temp            : ?? COMMA5.
        iwb             : ?? BEST1.
        wetb            : ?? COMMA5.
        dewpt           : ?? COMMA5.
        vappr           : ?? COMMA4.
        rhum            : ?? BEST3.
        msl             : ?? COMMA6.
        iwdsp           : ?? BEST1.
        wdsp            : ?? BEST2.
        iwddir          : ?? BEST1.
        wddir           : ?? BEST3.
        ww              : ?? BEST2.
        w               : ?? BEST2.
        sun             : ?? COMMA3.
        vis             : ?? BEST5.
        clht            : ?? BEST3.
        clamt           : ?? BEST1. ;
RUN;


data  wd.data_met;
set wd.data_met;
format Date_only date9.;
format time time.;

Date_only =datepart(date);
year =Year(Date_only);
month=month(Date_only);
week =week(Date_only);
year_month= catx("_",year,put(month,z2.));
```

```
year_week= catx("_",year,put(week(Date_only),z2.));
time = timepart(date);

label  rain = "Precipitation Amount (mm)";
label         temp ="Air Temperature (C)";
label         wetb ="Wet Bulb Temperature (C)";
label         dewpt =" Dew Point Temperature (C)";
label         vappr = "Vapour Pressure (hPa)";
label         rhum =  "Relative Hymidity (%)";
label         msl  = " Mean Sea Level Pressure (hPa)";
label         wdsp ="Mean Wind Speed (knot)";
label         wddir ="Predominant Wind Direction (degree)";
label         ww = "Synop code for Present Weather" ;
label         w  ="Synop code for Past Weather";
label         sun ="Sunshine duration (hours)" ;
label         vis ="Visibility (m)";
label         clht = "Cloud height - if none value is 999 (100's ft)" ;
label         clamt= "Cloud amount";

if time > "21:00:00"t or time < "6:00:00"t then delete;
run;




data wd.data_2015;
set wd.data_2015;
rename TOTEM_N11_STILLORGAN_RD__DATA_ON = TOTEM_N11_STILLORGAN_RD_;
rename TOTEM_ROCK_ROAD__DATA_ONLY_=TOTEM_ROCK_ROAD;
rename TOTEM_CLONSKEAGH_ROAD__DATA_ONLY = TOTEM_CLONSKEAGH_ROAD;

run;

%macro univariate_num(libin=,dsin=,libout=,dsout=);

*we use proc sql here to extract the base table metadata from dictionary.columns,
selecting all variables
 of type NUM that are not Dates;
      proc sql ;
            CREATE TABLE VAR_LIST as
                  select
                        upcase(name) AS VARIABLE
                        ,label
                        from dictionary.columns
                        where libname=upcase("&LIBIN.")
                              and memname=upcase("&DSIN.")
                              and type ="num"
                              and format ^="DATE9.";
      quit;

      *Count number of numeric vars we want to process;
      PROC SQL noprint;
            SELECT COUNT(VARIABLE) INTO :NVAR FROM VAR_LIST;
      QUIT;

      PROC SQL noprint;
            SELECT COUNT(*) INTO :NOBS FROM &LIBIN..&DSIN.;
      QUIT;

      *Create a series of macro variables containing the names of variables we
want to process;
      DATA _NULL_;
            length ii $4.;
            SET VAR_LIST end=last;
                  i+1;
                  ii=LEFT(PUT(i,4.));
                  call symputx('var'||ii, LEFT(VARIABLE));
```

```
                IF last THEN call symputx('NVAR', TRIM(LEFT(_N_)));
   RUN;

   *loop through all selected variables using proc means to claculate stats;
   %DO X=1 %TO &NVAR.;
   proc means data = &libin..&dsin. noprint nway missing;
    var &&VAR&X.;
output out = &&VAR&X.(drop=_:)
    n=num_populated
    nmiss = num_missing
    min = min_value
    max = max_value
    mean = avg_value
    std = st_deviation
    p10 = p10_value
    q1 = q1_value
    median = median_value
    q3 = q3_value
    p90 = p90_value
    ;
   run;
   %end;

   *stack stats together into one table;
   DATA stacked;
   SET
          %DO X = 1 %TO &NVAR.;
                &&VAR&X.
          %END;
   ;
   RUN;

   *merge with variable name and label;
   Data &libout..&dsout.;
   merge VAR_LIST stacked;
length flag $1 anomalous_reason $100;
   prop_missing = num_missing/&NOBS;

*logic to flag anomalous variables for deeper inspection;
if prop_missing = 1 then do;
    flag = 'A';
    anomalous_reason = '100% missing values';
end;
else if min_value = max_value then do;
    flag = 'A';
    anomalous_reason = 'All records take same value';
end;
else if prop_missing > 0.9 then do;
    flag = 'B';
    anomalous_reason = '>90% missing values';
end;
else if min_value = p90_value then do;
    flag = 'B';
    anomalous_reason = '>90% records take minimum value';
end;
else if max_value = p10_value then do;
    flag = 'B';
    anomalous_reason = '>90% records take maximum value';
end;
else if prop_missing > 0.75 then do;
    flag = 'C';
    anomalous_reason = '>75% missing values';
end;
else if p10_value = p90_value and p90_value ne . then do;
    flag = 'C';
    anomalous_reason = '>80% records take same value';
end;
else if min_value = q3_value then do;
```

```
        flag = 'C';
        anomalous_reason = '>75% records take minimum value';
    end;
    else if max_value = q1_value then do;
        flag = 'C';
        anomalous_reason = '>75% records take maximum value';
    end;
    else if prop_missing > 0.5 then do;
        flag = 'D';
        anomalous_reason = '>50% missing values';
    end;
    else if min_value = median_value then do;
        flag = 'D';
        anomalous_reason = '>50% records take minimum value';
    end;
    else if max_value = median_value then do;
        flag = 'D';
        anomalous_reason = '>50% records take maximum value';
    end;
    else if q1_value = q3_value and q3_value ne . then do;
        flag = 'D';
        anomalous_reason = '>50% records take same value';
    end;

      format sum_: min_: max_: avg_: st_: q1_: q3_: median_: p10_: p90_:
prop_missing: 3.2;
      run;

      *sort;
      proc sort data=&libout..&dsout.; BY DESCENDING prop_missing DESCENDING
anomalous_reason; RUN;

      *Delete intermediate tables stacked;
      PROC DATASETS LIBRARY= WORK;
      DELETE
          %DO X = 1 %TO &NVAR.;
                          &&VAR&X.
          %END;
             stacked
             VAR_LIST

          ;
      RUN;


%mend;

%univariate_num(libin=wd,dsin=data_2012,libout=WORK,dsout=UNIVAR_2012)
%univariate_num(libin=wd,dsin=data_2013,libout=WORK,dsout=UNIVAR_2013)
%univariate_num(libin=wd,dsin=data_2014,libout=WORK,dsout=UNIVAR_2014)
%univariate_num(libin=wd,dsin=data_2015,libout=WORK,dsout=UNIVAR_2015)

%macro inspect_meta(num_sheet=,);

%DO n=1 %TO &num_sheet.;
%Put *** Reading in &n. of &num_file. sheets;

proc sql;
      create table part_%eval(2011+&n.) as
            select
                  upcase(name) as variable_%eval(2011+&n.)
            from dictionary.columns
                  where libname ="WD"
                        AND memname="DATA_20%eval(11+&n.)";
quit;
%end;
```

```
	data unique_variables;
	set part_2012(rename=(variable_2012=variable))
		part_2013(rename=(variable_2013=variable))
		part_2014(rename=(variable_2014=variable))
		part_2015(rename=(variable_2015=variable)) ;
	run;

	proc sort data=unique_variables nodup; by variable; run;

	data unique_variables;
	set unique_variables;
	ID =_N_;
	run;


		PROC SQL;
		CREATE TABLE DQ_SUMAMRY AS
		SELECT t1.ID,
		t1.variable
		,case when t12.variable_2012 ="" then 0 else 1 end  as variable_2012
		,u12.prop_missing as prop_missing_2012
		,case when t13.variable_2013 ="" then 0 else 1 end  as variable_2013
		,u13.prop_missing as prop_missing_2013
		,case when t14.variable_2014 ="" then 0 else 1 end  as variable_2014
		,u14.prop_missing as prop_missing_2014
		,case when t15.variable_2015 ="" then 0 else 1 end  as variable_2015
		,u15.prop_missing as prop_missing_2015
		,case when calculated  variable_2012 + calculated variable_2013
				+ calculated variable_2014 +  calculated variable_2015 <
4 then 1 else 0 end as DQ_ISSUE_FLAG
		FROM WORK.UNIQUE_VARIABLES t1
		left join part_2012 t12 on t12.variable_2012=t1.variable
		left join UNIVAR_2012 u12 on u12.variable=t1.variable
		left join part_2013 t13 on t13.variable_2013=t1.variable
		left join UNIVAR_2013 u13 on u13.variable=t1.variable
		left join part_2014 t14 on t14.variable_2014=t1.variable
		left join UNIVAR_2014 u14 on u14.variable=t1.variable
		left join part_2015 t15 on t15.variable_2015=t1.variable
		left join UNIVAR_2015 u15 on u15.variable=t1.variable
		 ;
QUIT;

data DQ_SUMAMRY_FINAL;
retain ID variable label DQ_ISSUE_FLAG variable_2012 prop_missing_2012
variable_2013 prop_missing_2013 variable_2014 prop_missing_2014 variable_2015
prop_missing_2015 ;
set DQ_SUMAMRY;
format label $char100.;

if variable ="BICYCLE_IN" then label = "Location 2: Cyclists in ";
if variable ="BICYCLE_OUT" then label = "Location 2: Cyclists out";
if variable ="WALKING_IN" then label = "Location 2: Pedestrians in ";
if variable ="WALKING_OUT" then label = "Location 2: Pedestrians out";
if variable ="GLENAGEARY" then label = "Location 2: Total of walking in/out +
cyclists in/out";
if variable ="IN__TOWARDS_CITY_CENTRE" then label = "Location 3: cyclists in ";
if variable ="OUT_AWAY_FROM_CITY_CENTRE" then label = "Location 3: cyclists out";
if variable ="N11_MONTROSE" then label = "Location 3: Total of cyclists in/out";
if variable ="ROCK_ROAD__BUS_LANE_BESIDE_PARK" then label = "Location 4: Cyclists
1-way ";
if variable ="ROCK_ROAD_PARK_IN" then label = "Location 5: Cyclists in ";
if variable ="ROCK_ROAD_PARK_OUT" then label = "Location 5: Cyclists out";
if variable ="ROCK_ROAD__PARK" then label = "Location 5: Total cyclists in/out ";
if variable ="DATE" then label = "Date";
if variable ="N11_ECO_TOTEM" then label = "Unkown";
if variable ="ROCK_ROAD_INBOUND" then label = "Unkown";
if variable = "TOTEM_CLONSKEAGH_ROAD" then label ="Location 6: Total of cyclists
in/out";
```

```
if variable = "TOTEM_IN" then label ="Location 6: Cyclists in";
if variable = "TOTEM_OUT" then label ="Location 6: Cyclists out";
if variable = "TOTEM_N11_STILLORGAN_RD_" then label ="Location 1: Cyclists 1-way";
if variable = "TOTEM_ROCK_ROAD" then label ="Location 7: Cyclists 1-way";

run;

proc sort data=dq_sumamry_final; by label; run;


%mend;
%inspect_meta(num_sheet=4);

proc sql;
	select
		trim(cat("label ",trim(variable)," = ","'",trim(label),"'",";"))
	into :label_desc SEPARATED  by " "
		from
			dq_sumamry_final
	;
quit;


data wd.all_counter_data;
set wd.data_2012 wd.data_2013 wd.data_2014 wd.data_2015;

%put &label_desc.

year =Year(date);
month=month(date);
week =week(date);
year_month= catx("_",year,put(month,z2.));
year_week= catx("_",year,put(week(date),z2.));
run;



%macro histplot(dsin=,yaxis=,title=);
%_sas_pushchartsize(1000,600);

Legend1 FRAME CBORDER=BLACK CFRAME=LTGRAY POSITION = (TOP RIGHT INSIDE)
LABEL=("Direction: ") MODE=PROTECT;
Axis1 STYLE=1 WIDTH=1 MINOR=NONE LABEL=(&yaxis.);
Axis2 STYLE=1 WIDTH=1 LABEL=("Year Month");
Axis3 LABEL=NONE;
TITLE;
TITLE1 &title.;
/*FOOTNOTE;*/
PROC GCHART DATA= &dsin.
;
	VBAR
	 year_month
 /
	SUMVAR=Volume1
	SUBGROUP=Source
	GROUP=year
	CLIPREF
FRAME	CFRAME=LTGRAY
	TYPE=SUM
	NOZERO
	LEGEND=LEGEND1
	COUTLINE=BLACK
	RAXIS=AXIS1
	MAXIS=AXIS2
	GAXIS=AXIS3
	LREF=4
	CREF=GRAY
```

```
       AUTOREF
;
RUN; QUIT;
%mend histplot;

%macro tab(dsin=,total_var=);
TITLE;
TITLE1;
FOOTNOTE;
PROC TABULATE
DATA=&dsin.
       FORMAT=COMMA12.
       ;

       VAR &total_var.;
       CLASS Year /  ORDER=UNFORMATTED MISSING;
       CLASS Month / ORDER=UNFORMATTED MISSING;
       TABLE
       Year,Month*Sum={LABEL="" STYLE={NOBREAKSPACE=ON}}*
    &total_var.={LABEL="" STYLE={NOBREAKSPACE=ON}}
/ INDENT=0          ;
       ;
RUN;
%mend tab;


/*Location 1: N11 Stillorgan Road Totem (1-way)*/

PROC SQL;
   CREATE TABLE location_1 AS
   SELECT     year_month
                  ,put(year,4.) as Year
                  ,max(date) as Month format monname3.
                  ,coalesce(SUM(totem_n11_stillorgan_rd_),0) AS TOTAL_1_WAY
      FROM WD.ALL_COUNTER_DATA
      GROUP BY
                  1,2;
QUIT;


PROC TRANSPOSE DATA=location_1
       OUT=location_1_trans
       PREFIX=Volume
       NAME=Source
       LABEL=Label
;
       BY year_month year;
       VAR  TOTAL_1_WAY   ;;

RUN; QUIT;

%histplot(dsin=location_1_trans,yaxis="Total Cyclists (1-way)",title="Location 1:
N11 Stillorgan Road Totem (1-way)");
%tab(dsin=location_1,total_var=TOTAL_1_WAY);


/*Location 2: The Metals, Glenageary Dart Station (2-way cyclists with pedestrians
excluded)*/

PROC SQL;
   CREATE TABLE location_2 AS
   SELECT
            year_month
            ,put(year,4.) as year
            ,max(date) as Month format monname3.
            ,coalesce(SUM(BICYCLE_IN),0) AS BICYCLE_IN
            ,coalesce(SUM(BICYCLE_OUT),0) AS BICYCLE_OUT
            ,calculated BICYCLE_IN +calculated BICYCLE_out as Total
```

```
              ,coalesce(SUM(WALKING_IN),0) AS WALKING_IN
              ,coalesce(SUM(WALKING_OUT),0) AS WALKING_OUT
       FROM WD.ALL_COUNTER_DATA
       GROUP BY
                     1,2;
QUIT;

PROC TRANSPOSE DATA=location_2
       OUT=location_2_trans
       PREFIX=Volume
       NAME=Source
       LABEL=Label
;
       BY year_month year;
       VAR  BICYCLE_IN BICYCLE_OUT  ;;

RUN; QUIT;

%histplot(dsin=location_2_trans,yaxis="Total Cyclists (2-way)",title="Location 2:
The Metals, Glenageary Dart Station - (2-way cyclists)");
%tab(dsin=location_2,total_var=Total);

/*Location 3: N11 Montrose (2-way cycle track)*/

PROC SQL;
   CREATE TABLE location_3 AS
   SELECT
              year_month
              ,put(year,4.) as year
              ,max(date) as Month format monname3.
              ,coalesce(SUM(IN__TOWARDS_CITY_CENTRE),0) AS BICYCLE_IN
              ,coalesce(SUM(OUT_AWAY_FROM_CITY_CENTRE),0) AS BICYCLE_OUT
              ,calculated BICYCLE_IN +calculated BICYCLE_out as Total
       FROM WD.ALL_COUNTER_DATA
       GROUP BY
                     1,2;
QUIT;



PROC TRANSPOSE DATA=location_3
       OUT=location_3_trans
       PREFIX=Volume
       NAME=Source
       LABEL=Label
;
       BY year_month year;
       VAR  BICYCLE_IN BICYCLE_OUT  ;;

RUN; QUIT;

%histplot(dsin=location_3_trans,yaxis="Total Cyclists (2-way)",title="Location 3:
N11 Montrose (2-way cycle track)");
%tab(dsin=location_3,total_var=Total);


/*Location 4:Rock Road Bus lane  (1-way)*/

PROC SQL;
   CREATE TABLE location_4 AS
   SELECT
              year_month
              ,put(year,4.) as year
              ,max(date) as Month format monname3.
              ,coalesce(SUM(ROCK_ROAD__BUS_LANE_BESIDE_PARK),0) AS TOTAL_1_WAY
       FROM WD.ALL_COUNTER_DATA
       GROUP BY
                     1,2;
```

```
QUIT;

PROC TRANSPOSE DATA=location_4
       OUT=location_4_trans
       PREFIX=Volume
       NAME=Source
       LABEL=Label
;
       BY year_month year;
       VAR  TOTAL_1_WAY  ;;

RUN; QUIT;

%histplot(dsin=location_4_trans,yaxis="Total Cyclists (1-way)",title="Location 4:
Rock Road Bus lane  (1-way)");
%tab(dsin=location_4,total_var=TOTAL_1_WAY);

/*Location 5:Rock Road Inside park near the Dart (2-way)*/

PROC SQL;
   CREATE TABLE location_5 AS
   SELECT
              year_month
              ,put(year,4.) as year
              ,max(date) as Month format monname3.
              ,coalesce(SUM(ROCK_ROAD_PARK_IN),0) AS BICYCLE_IN
              ,coalesce(SUM(ROCK_ROAD_PARK_OUT),0) AS BICYCLE_OUT
              ,calculated BICYCLE_IN +calculated BICYCLE_out as Total

      FROM WD.ALL_COUNTER_DATA
      GROUP BY
                    1,2;
QUIT;

PROC TRANSPOSE DATA=location_5
       OUT=location_5_trans
       PREFIX=Volume
       NAME=Source
       LABEL=Label
;
       BY year_month year;
       VAR  BICYCLE_IN BICYCLE_OUT   ;;

RUN; QUIT;

%histplot(dsin=location_5_trans,yaxis="Total Cyclists (2-way)",title="Location
5:Rock Road Inside park near the Dart (2-way)");
%tab(dsin=location_5,total_var=Total);

/*Location 6: Clonskeagh Road Totem - (2-way)*/

PROC SQL;
   CREATE TABLE location_6 AS
   SELECT
              year_month
              ,put(year,4.) as year
              ,max(date) as Month format monname3.
              ,coalesce(SUM(TOTEM_IN),0) AS BICYCLE_IN
              ,coalesce(SUM(TOTEM_OUT),0) AS BICYCLE_OUT
              ,calculated BICYCLE_IN +calculated BICYCLE_out as Total
      FROM WD.ALL_COUNTER_DATA
      GROUP BY
                    1,2;
QUIT;

PROC TRANSPOSE DATA=location_6
       OUT=location_6_trans
       PREFIX=Volume
```

```
       NAME=Source
       LABEL=Label
;
       BY year_month year;
       VAR  BICYCLE_IN BICYCLE_OUT   ;;

RUN; QUIT;

%histplot(dsin=location_6_trans,yaxis="Total Cyclists (2-way)",title="Location 6:
Clonskeagh Road Totem - (2-way)");
%tab(dsin=location_6,total_var=Total);

/*Location 7: Rock Road Totem  (1-way)*/

PROC SQL;
   CREATE TABLE location_7 AS
   SELECT
             year_month
             ,put(year,4.) as year
             ,max(date) as Month format monname3.
             ,coalesce(SUM(TOTEM_ROCK_ROAD),0) AS TOTAL_1_WAY

      FROM WD.ALL_COUNTER_DATA
      GROUP BY
                   1,2;
QUIT;

PROC TRANSPOSE DATA=location_7
       OUT=location_7_trans
       PREFIX=Volume
       NAME=Source
       LABEL=Label
;
       BY year_month year;
       VAR  TOTAL_1_WAY    ;;

RUN; QUIT;

%histplot(dsin=location_7_trans,yaxis="Total Cyclists (1-way)",title="Location 7:
Rock Road Totem  (1-way)");
%tab(dsin=location_7,total_var=TOTAL_1_WAY);
```