Detecting Journeys in Bicycle Sharing Systems from Docking Station Counts

James Todd, Oliver O'Brien and James Cheshire

Department of Geography, University College London

January 31, 2019

Summary

Changes in numbers of available bicycles at bikeshare docking stations in individual cities were counted to detect journeys commencing. These counts were then compared with published journey records, with an error factor calculated. Similar cities, without journey data available, were identified, and the count data and error factors were combined to predict the numbers of journeys occurring. The method found reasonable success, with the great majority of journeys detected, and the potential to expand the technique to many other cities for which only dock count data is available.

KEYWORDS: bicycle sharing systems, docking stations, bicycle journeys, MAAS, micro-mobility

1. Introduction

Dock-based bicycle sharing systems have seen significant growth, particularly over the last ten years. They provide an easy way to allow tourists, commuters and other utility/leisure users to travel within an urban area, between docking stations at fixed locations.

These bicycle sharing systems lend themselves to urban mobility research since many publish feeds of near real-time data on the numbers of bicycles available in each of their docking stations. The feeds provide bike availability data for each docking station. We believe it is possible to infer journeys based on the small fluctuations that occur throughout the day. These estimates can then be validated against known journeys taken since some systems release these data periodically in batches. Whilst the proposed approach may lack precision, it offers the potential for the real-time monitoring of journeys, as well as estimates for systems that do not record or release full journey information

Quantitative analysis of docking station data, to understand spatiotemporal variations in how individual bicycle sharing systems are used, has been carried out by a number of papers, including O'Brien (2014) which classified city-based systems based on intraday usage peaks and spatial arrangement and density of docking stations. Chardon (2015) looked at docking station data to estimate journeys. This applied a number of models and were able to closely predict journey counts. This paper aims to extend on the techniques used by Chardon in order to accurately predict the number of journeys within a system for which there is no journey level data, by calibrating with a similar system.

2. Method

Data were collected from public web services every 2 minutes. These data contain simple counts of bikes within each dock and the timestamp for which the data was collected. Data collection commenced, for some cities such as London, as far back as late 2010, and over 400 cities how have the data collected. This paper utilises the most recent data, mainly investigating trends within 2016-2018 for each city. The journey data for the cities with this data available were downloaded from similar repositories and aggregated to daily counts.

We chose four city pairs: a calibration city with both dock-based data and journey data, and a prediction city where only the first type of data is available, for which we would predict the number of journeys within the system. It was, therefore, necessary to pair these cities based on similar characteristics to ensure a justified prediction. Table 1 shows the characteristics which were chosen to match cities, with size, average daily maximum simultaneous usage in January and July 2018, and average temperatures for these months (CustomWeather, 2018) of the systems considered.

Table 1 Classification and pairing of city bikeshare systems based on size, use and use variability, climate and available journey data. Use variability was the principal component used to pair the cities.

City	Public Journey Data	January 2018			July 2018		
		Max Bikes Avail.	Max Simult. Use %	Average Temp.	Max Bikes Avail.	Max Simult. Use %	Average Temp.
London	Yes	10051	10%	6°C	8940	19%	21°C
Paris (2017)	No	18901	11%	2°C	14493	19%	20°C
Madrid	Yes	1893	12%	8°C	1732	16%	28°C
Barcelona	No	4955	19%	12°C	4374	24%	26°C
Chicago	Yes	3345	5%	-3°C	3888	24%	25°C
Toronto	No	2219	4%	-5°C	2511	21%	24°C
Montreal	Yes	0	0%	-9°C	5486	20%	24°C
Moscow	No	0	0%	-4°C	3841	30%	20°C

The climate and temperature within these cities play a significant role in cycling levels and by extension the use of bicycle sharing systems (Miranda-Moreno, 2010). “Cold” cities, such as Montreal and Moscow, which experience very cold winters, were matched together, as user behavior was likely to be similar. “Cold” and “warm” cities, when operating, show very different trends in usage and user type (OBIS, 2011).

In order to detect the number of journeys from the dock-level data, it was necessary to aggregate the changes in the number of bikes within in a particular dock chronologically. Data for each dock was extracted from the database and ordered by time. Comparing a row to its previous row enabled us to detect changes in the number of bikes, from which we inferred that each single decrease was the start of a journey. This data was combined with the equivalent from all docking to calculate the estimated daily total number of journeys within the bike sharing system.

Data from each of the calibration cities were combined with separately published aggregated journey data. The accuracy of the dock-based journey estimates were calculated as the percentage difference “error” between the two sources. These were then applied to each respective prediction city, the quarterly numbers for which can be found within Appendix 1. The percentage differences were then applied, on an annually aggregated basis, to the estimated journey numbers for those prediction cities. These numbers can be found in Appendix 2.

3. Results

The following results report the journey estimations from dock-based data, as well as the journey counts for those calibration cities. Looking at the calibration cities, we can gain an understanding of how well the docking station data fits the journey data for the respective city. We find both overcounts and undercounts within our results, for which there are several explanations.

Undercounts may be due to docking and undocking activity happening at individual docking stations within the interval used to observe the number of bicycles available (this interval is typically 2 minutes for the systems studied). Commuting flows during “rush-hours” are where many systems see their peak usage, these flows are typically unidirectional and therefore less susceptible to this undercount, except where the operator manually refills (or empties) certain docking stations manually, during peak journey times, to extend the capacity of individual docking stations.

Conversely, overcounts may be as a result of rebalancing activity being detected, but not being listed in the corresponding journey datasets (a rebalancing removal, or a retrieval of a bicycle for maintenance by the operator, is not a customer journey and so reasonably could not be included in some cities’ datasets). The proportion of “journeys” being rebalancing operations can be significant – for example, in New York City, approximately 6% of journeys are rebalancing (Teigland 2015). Overcounts may also be caused by data feed errors (e.g. electrical connection noise), resulting in phantom journey starts.

It is necessary to analyse weekday and weekend journey counts separately as these typically vary in volume very significantly. For brevity we present just the weekday results in this section.

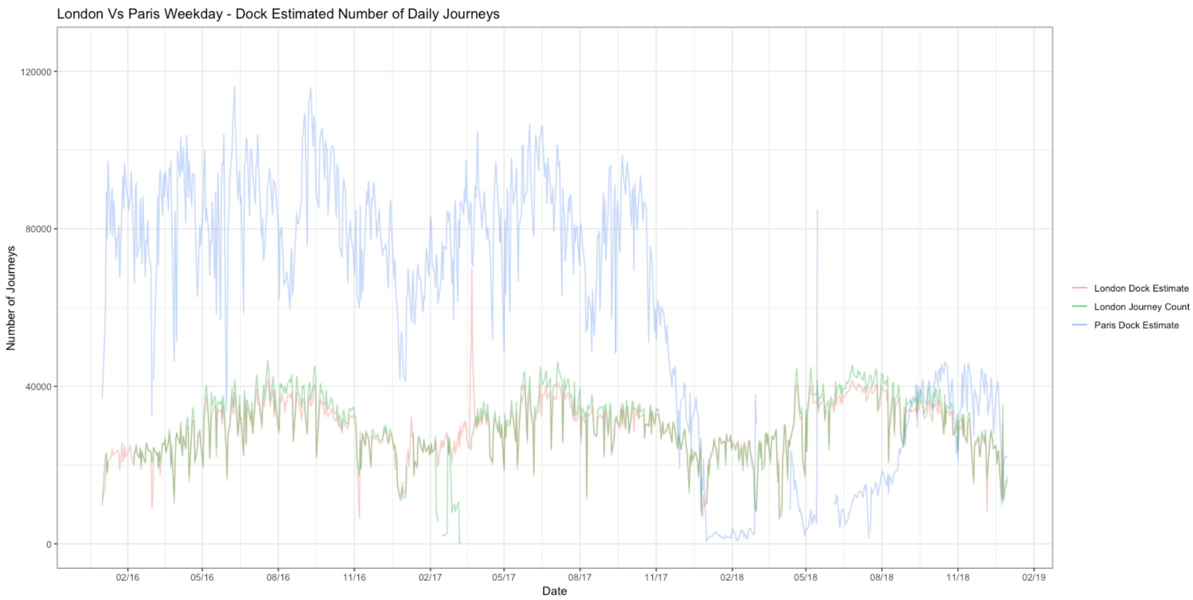

Figure 1 Weekday Journey Counts for London and Estimated Journeys for London and Paris.

Figure 1 shows that the dock-based journey estimates in London match well with the known journey data. The difference between the two data sources is between -10% and +32%, although for the majority of the measurement periods, we find there to be a slight underestimation of journeys. Paris sees a significant decrease in the beginning of 2018 - we can attribute this to a change in operator which disrupted operations significantly (Chrisafis, 2018). Our dock data suggests a 50% fall in the journey count from 2017 to 2018.

Figure 2 Weekday Journey Counts for Madrid and Estimated Journeys for Madrid and Barcelona.

For Madrid, the published journeys and estimates appear to closely match each other throughout the period (Figure 2). The dock estimates are a slight underestimation of the journey counts data, ranging between -3% and -11%. With Barcelona appearing to have similar characteristics to Madrid, the dock data for Barcelona is likely to similarly be a good predictor of journey counts.

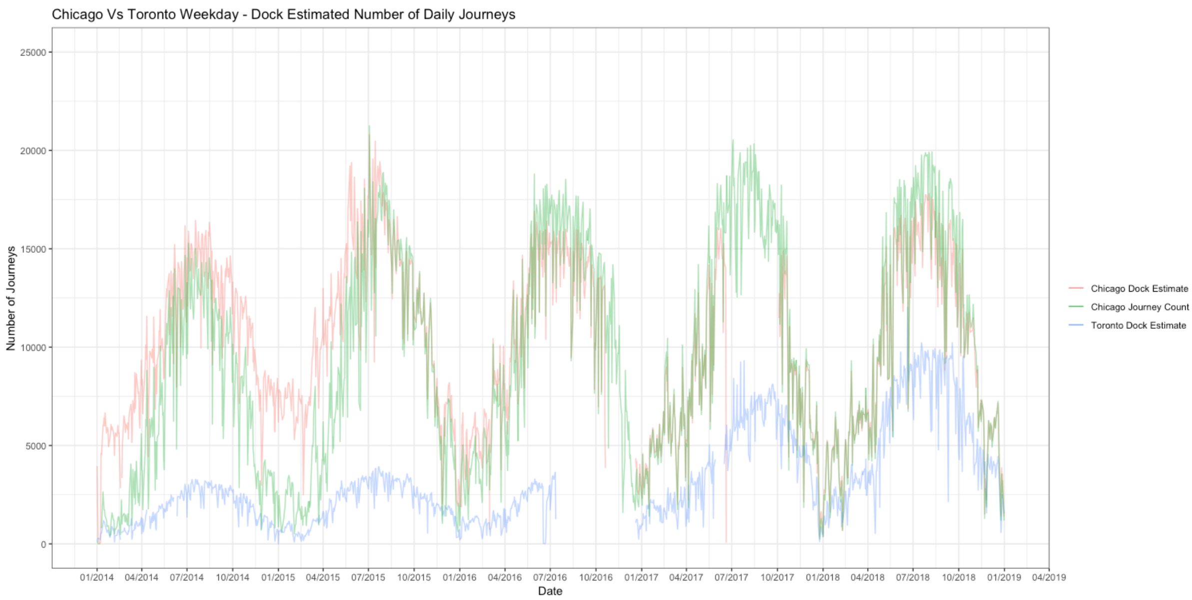

Figure 3 Weekday Journey Counts for Chicago and Estimated Journeys for Chicago and Toronto.

Journey variations with season, for Chicago (Figure 4), are very significant, and changing user types and journey frequencies may be the cause of over and under-estimations in our results ranging between -19% and +83%. This may cause issues when estimating journeys for Toronto. There are also some large overestimations for in some winter months, which lead us to believe that there are issues with the data within these periods, or the system is being used in winter-time in these "cold" cities in an unusual way.

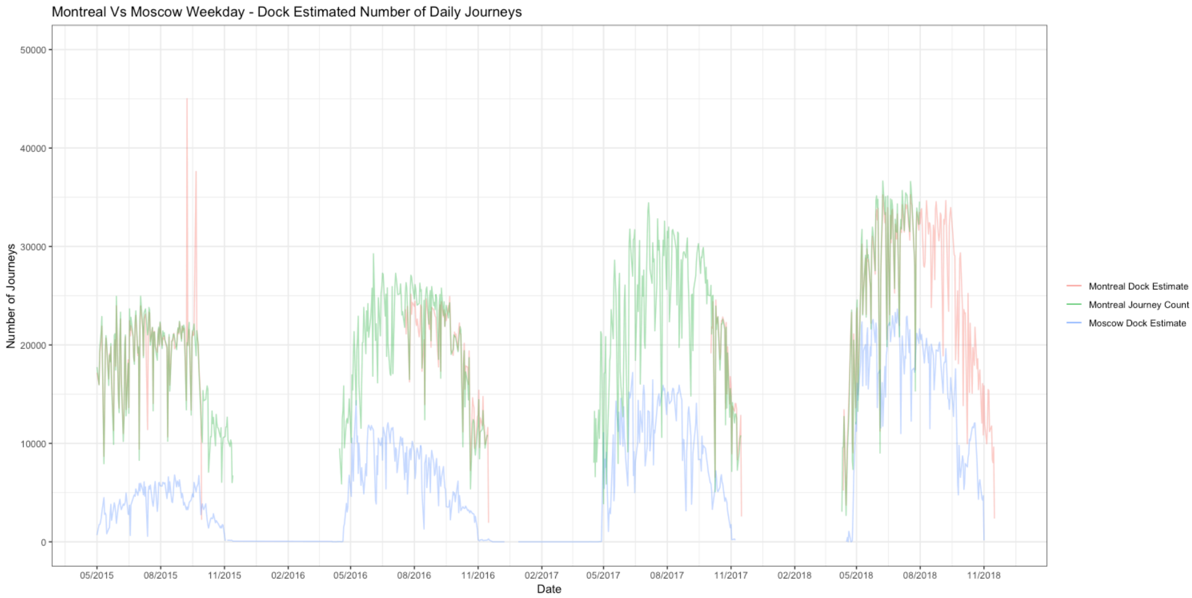

Figure 4 Weekday Journey Counts for Montreal and Estimated Journeys for Montreal and Moscow.

In Montreal and Moscow, also "cold weather" cities, both bicycle sharing systems shut down over the winter months, which we can clearly see within Figure 4. When the Montreal is operating, there appears to be a very close correlation between dock and journey data, with the error varying from -8% to +13%. There does not seem to be a consistent over or underestimation, but results suggest that the docking station data for Montreal provides a good prediction for journey counts there and so likely for the similar system in Moscow.

To validate the predicted journey counts, we examined full-year summary statistics, which are typically published by press releases. Paris reported 38.1 million journeys in 2016 (Elbaz-Baratti, 2017). Our system detected 27.4 million journeys (Table 6), with an expected 2-10% of journeys not detected if the system is used in a similar way to London's (Table 2). Moscow reported 4.25 million journeys in 2018 (Sobyanin, 2018). Our system detected 3.34 million journeys (Table 9) with our paired calibration city (Montreal) suggesting little adjustment for undetected journeys is required. It is not known how the "headline" reported journey number is calculated.

4. Discussion

The results suggest both that dock data is for many cities a good proxy for journey information and that if cities are paired based on operational similarity (e.g. climate characteristics, size and commute/leisure usage volumes), then journeys can be inferred with reasonably quantitative accuracy in the absence of granular city statistics (e.g. daily journey counts).

5. Conclusions

This paper investigates the use of dock-based data to estimate journeys within a particular bicycle sharing system. The results we have seen are very promising, with many of the pairings exemplifying the validity of our methodology. With further tuning and adaptation, this method may be used as a method to measure journeys accurately, a metric which is hard to come by and a useful in measure of the success of a bicycle sharing system.

Acknowledgements

This research is funded by the Economic and Social Research Council (ESRC). Docking station data was supplied by CDRC Data.

The authors would like to thank the GLA and TfL for accelerated delivery of recent journeys data for London, and to all the bikeshare operators and transport authorities for the cities included in this research for their open supply of the dock data and where available journey data.

References

Chardon M., Caruso G. (2015) Estimating bike-share trips using station level data. *Transportation Research Part B: Methodological* **78**(C) 260-279. doi:10.1016/j.trb.2015.05.003

Chrisafis, A. (2018) Wheels come off Paris bike-share scheme after hi-tech upgrade. *The Guardian.* https://www.theguardian.com/world/2018/may/04/paris-bike-share-scheme-velib-hi-tech-upgrade-problems.

CustomWeather (2018) Weather records data, https://www.timeanddate.com/weather/

Elbaz-Baratti, C. (2017) Velib' Metropole. Velocittà conference, 16 November 2017. http://velo-citta.eu/wp-content/uploads/20171115-CBE-Velib-Paris.pdf.

Miranda-Moreno L., Nosal T. (2010) Weather or Not to Cycle: Temporal Trends and Impact of Weather on Cycling in an Urban Environment. *Transportation Research Record* **2247**(1) 42-52. doi:10.3141/2247-06

OBIS (2011) Optimising Bike Sharing in European Cities – A Handbook pp31-32. *Intelligent Energy Europe Proramme.* http://mobility-workspace.eu/wp-content/uploads/OBIS_Handbook_EN.pdf.

O'Brien O., Cheshire J., Batty M. (2014) Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography* **34** 262-273. doi:10.1016/j.jtrangeo.2013.06.007

Sobyanin., S. (2018) https://twitter.com/MosSobyanin/status/1058362160848482312.

Teilgland., N. (2015) Finding Balance: Estimating bikeshare rebalancing events using trip data. *ESRI User Conference 2015.* http://proceedings.esri.com/library/userconf/muc15/papers/muc_17.pdf

Biographies

James Todd is currently a first year PhD student funded by the UBEL DTP in collaboration with Arup. His PhD will investigate the utility of new urban data, in areas such as transportation and the sharing economy.

Oliver O'Brien is an urban data researcher, data scientist and software developer at the Consumer Data Research Centre at UCL, where he specialises in visualising city datasets. He has created a number of software products including Bike Share Map, CityDashboard, DataShine, CensusProfiler, Named, and CDRC Maps. He edits Mapping London.

James Cheshire is an associate professor at UCL Geography and director of UCL QStep. He is president of the Society of Cartographers and co-author of Where the Animals Go and London: The Information Capital.

Appendix 1: Quarterly Journey Estimation Tables

Table 2 Average Daily Journeys in London and Paris

Quarter	London Journey Counts		London Journey Predictions		Paris Journey Predictions	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Q1 2016	22,646	13,299	22,222 (-2%)	14,791 (10%)	79,056	59,132
Q2 2016	31,002	28,689	29,649 (-5%)	26,957 (-6%)	87,631	67,868
Q3 2016	37,996	34,436	35,257 (-8%)	31,349 (-10%)	88,451	72,367
Q4 2016	27,561	19,467	26,454 (-4%)	19,418 (0%)	78,116	59,804
Q1 2017	18,232	11,979	26,977 (32%)	16,450 (27%)	74,309	54,342
Q2 2017	33,656	31,705	31,864 (-6%)	29,335 (-8%)	87,081	72,649
Q3 2017	34,208	29,843	32,417 (-6%)	27,825 (-7%)	79,541	65,446
Q4 2017	28,463	18,006	27,157 (-5%)	17,652 (-2%)	52,242	38,076
Q1 2018	22,518	13,533	23,174 (3%)	14,871 (9%)	3,530	2,059
Q2 2018	35,035	30,941	33,413 (-5%)	28,917 (-7%)	92,961	76,557
Q3 2018	38,025	30,762	35,646 (-7%)	28,691 (-7%)	23,322	21,018
Q4 2018	28,075	17,262	27,094 (-4%)	17,619 (2%)	36,729	27,165

Table 3 Average Daily Journeys in Madrid and Barcelona

Quarter	Madrid Journey Counts		Madrid Journey Predictions		Barcelona Journey Predictions	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Q2 2017	11,572	8,534	11,133 (-4%)	8,065 (-6%)	42,587	28,232
Q3 2017	12,640	8,938	11,581 (-9%)	8,449 (-6%)	36,387	24,937
Q4 2017	11,834	9,099	10,668 (-11%)	8,475 (-7%)	36,349	23,362
Q1 2018	8,649	5,964	7,787 (-11%)	5,763 (-3%)	34,066	20,789
Q2 2018	11,511	8,710	10,667 (-8%)	8,011 (-9%)	39,320	23,665
Q3 2018	13,223	8,880	11,879 (-11%)	8,152 (-9%)	35,754	22,698

Table 4 Average Daily Journeys in Chicago and Toronto

Quarter	Chicago Journey Counts		Chicago Journey Predictions		Toronto Journey Predictions	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Q1 2014	1,591	915	5,598 (72%)	5,341 (83%)	675	386
Q2 2014	7,880	10,424	10,770 (27%)	12,414 (16%)	4,183	2,469
Q3 2014	11,790	12,801	13,948 (15%)	14,410 (11%)	2,775	1,989
Q4 2014	5,093	3,917	94,277 (46%)	8,159 (52%)	1,534	878
Q1 2015	2,491	1,650	76,76 (68%)	7,021 (76%)	759	399
Q2 2015	9,623	10,284	13,732 (30%)	13,738 (25%)	2,493	1,727
Q3 2015	15,314	17,025	15,489 (1%)	15,585 (-9%)	3,200	2,404
Q4 2015	7,465	5,297	8,164 (8%)	6,044 (12%)	1,867	1,041
Q1 2016	4,795	3,268	5,943 (19%)	4,779 (32%)	1,054	639
Q2 2016	11,789	11,748	12,028 (2%)	11,502 (-2%)	2,232	1,510
Q3 2016	15,258	16,659	13,940 (-9%)	14,541 (-15%)	2,885	2,876
Q4 2016	7,967	613	8,086 (1%)	6,863 (11%)	917	463
Q1 2017	5,230	3,668	5,300 (1%)	3,873 (5%)	1,827	992
Q2 2017	12,296	12,328	10,586 (-16%)	10,362 (-19%)	3,526	2,734
Q3 2017	17,364	17,762	-	-	6,891	5,826
Q4 2017	8,271	4,875	7,741 (-7%)	4,451 (-10%)	4,461	2,685
Q1 2018	4,840	2,901	4,892 (1%)	3,322 (13%)	2,426	1,629
Q2 2018	12,327	9,939	11,583 (-6%)	9,340 (-6%)	6,521	4,750
Q3 2018	16,980	15,179	15,263 (-11%)	13,697 (-11%)	8,793	7,550
Q4 2018	8,064	4,249	7,765 (-4%)	4,347 (2%)	4,608	2,743

Table 5 Average Daily Journeys in Montreal and Moscow

Quarter	Montreal Journey Counts		Montreal Journey Predictions		Moscow Journey Predictions	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Q2 2015	18,384	16,801	18,262 (-1%)	16,895 (1%)	3,243	3,466
Q3 2015	19,878	16,529	20,380 (4%)	16,275 (-2%)	4,985	5,828
Q4 2015	11,354	6,835	-	-	1,623	1,674
Q1 2016	-	-	-	-	-	-
Q2 2016	19,060	16,609	-	-	6,768	7,560
Q3 2016	23,571	18,501	22,102 (-7%)	18,322 (-1%)	7,254	8,346
Q4 2016	13,692	7,965	13,842 (1%)	92,19 (13%)	1,773	2,139
Q1 2017	-	-	-	-	-	-
Q2 2017	20,115	19,453	-	-	7,710	10,845
Q3 2017	27,899	24,853	-	-	11,953	12,891
Q4 2017	17,041	12,255	16,772 (-2%)	12,612 (3%)	4,092	3,219
Q1 2018	-	-	-	-	-	-
Q2 2018	24,215	22,117	24,648 (2%)	21,940 (-1%)	15,338	17,307
Q3 2018	31,800	28,150	30,327 (-4%)	26,001 (-8%)	17,227	18,913
Q4 2018	-	-	15,728 (-)	11,212 (-)	7,511	7,995

Appendix 2: Prediction City Annual Journal Predictions

Table 6 Total Annual Journey Prediction for Paris

Year	Paris Weekday		Paris Weekend		Prediction
	Total Journeys	Average Error	Total Journeys	Average Error	with Error Applied
2016	21,754,260	-4.75%	6,798,245	-1.5%	27,400,000
2017	19,056,197	3.75%	6,042,517	2.5%	26,000,000
2018	10,504,821	-3.25%	3,467,638	-0.75%	13,600,000

Table 7 Total Annual Journey Prediction for Barcelona

Year	Barcelona Weekday		Barcelona Weekend		Prediction
	Total Journeys	Average Error	Total Journeys	Average Error	with Error Applied
2018	7,499,421	-10.25%	1,869,033	-7%	8,470,000

Table 8 Total Annual Journey Prediction for Toronto

Year	Toronto Weekday		Toronto Weekend		Prediction
	Total Journeys	Average Error	Total Journeys	Average Error	with Error Applied
2014	593,447	40%	147,755	40.5%	10,400,000
2015	545,012	26.75%	145,843	26%	875,000
2016	434,810	3.25%	121,104	6.5%	578,000
2017	1,094,727	-7.3%	327,605	-7%	1,320,000
2018	1,457,212	-5%	439,396	-0.5%	1,820,000

Table 9 Total Annual Journey Prediction for Moscow

Year	Moscow Weekday		Moscow Weekend		Prediction
	Total Journeys	Average Error	Total Journeys	Average Error	with Error Applied
2016	988,669	-3%	430,534	6%	1,420,000
2017	1,394,866	-2%	652,649	3%	2,040,000
2018	2,374,188	-1%	1,039,180	-4.5%	3,340,000