Capstone: City of Melbourne Parking

Trevor Alback: General Assembly DAi 2021

Contents

Part 2: Report and Technical Analysis

Introduction

This document details a report and technical analysis for the client brief given on parking in the city of Melbourne. The client in this case is the Melbourne City Council (MCC), which is the government body responsible for the municipality of Melbourne. The area in question includes the CBD and surrounding 37 square kilometers.

The role of the MCC is to plan and provide services, facilities, and infrastructure for the community. Parking and transport fall within these areas of responsibility. This project investigates and gives recommendations specifically related to on-street car parking bays (i.e. not private or off-street parking).

The report is best read with reference to the provided technical documentation. For example, Api information is provided within Juypyter notebooks so that up-to-date data through the city of Melbourne's Open data source can be accessed. The notebooks also contain coding with comments demonstrating the steps from data calling and cleaning to analysis and saving cleaned or aggregated files. Exported Microsoft Excel files and Tableau workbooks show visualizations of the cleaned data. A summary of the submitted files is provided in the appendix for reference.

Problem statements and potential benefits

The two problem statements given by the client to investigate are:

- Removal of some on-street parking within the city limits
- Optimize park bay restrictions to minimize breaches across the city

Improvements related to the two statements can make city areas more amenable to public transport and pedestrians, increase effective use of existing spaces and help to manage traffic flows and congestion. The MCC already has a 10-year transport strategy to 2030. It focuses on a long-term vision for city commuters, with goals to give more space to pedestrians, cyclists, and for greening. Melbourne's daily population is projected to increase to 1.4 million by 2036 and repurposing parking spaces is a key initiative to manage this increase (MCC, 2021).

Data Sources, definitions, and data dictionary

The city of Melbourne provide access to some of their data through an open data platform. An Api endpoint is given so that relevant data can be called directly into python workbooks. Api's have the advantage of bringing up-to-date data into notebooks for analysis. The city of Melbourne website provides details for how to access data and the process is shown in the Juypter notebooks (MCC, 2021). It is best to get a free app token to use for Api calls so throttling doesn't occur.

The analysis in this report focuses on the provided 2017 on-street parking sensor data (MCC, 2021). This file has 35.9 million rows and is 5.4gb. It is a large file that is hard to handle efficiently as a csv download. It also means analysis of the whole dataset through the Api is not realistic due to the file size. To get around this, filters can be used when calling the data through the api, for example, by area, as is shown in the python notebooks. Each row in this dataset is a unique parking instance. The data set is clean with only two columns containing some nulls. These columns with nulls are not used in the analysis process.

A second data set, On-street Parking Bay spatial polygons, is used in the analysis because it has accessible latitude and longitude coordinate points for parking bays. This data can be called through the Api. Each dataset has a unique identifier available on the open data webpage and shown in the notebooks. Other datasets provided, but not used in this analysis were a restrictions data set, giving more detailed information of parking bay restrictions and a live parking information dataset. Links for all datasets can be found in the references section. The data dictionary is below.

Column Name	Description
2017 On-street parking sensor data (often referred to as sensor_df)	
DeviceId	Serial number of the in-ground sensor
ArrivalTime	Date and time that sensor detects a vehicle over it
DepartureTime	Date and time that sensor detects a vehicle no longer over it
DurationSeconds	The difference between arrival and departure events (in seconds)
StreetMarker	The street marker located next to parking bay with a unique id. This is the key to connect to the location_df
Sign	Description/restriction on parking sign at time of parking event
Area	City area
StreetName	Street where the vehicle is parked
In Violation	True = In violation, False = Not in violation
On-street parking Bay polygons (often referred to as location_df)	
The_geom	Multipolygon coordinate information. Information is extracted from this column to get a latitude and longitude parking bay coordinate
Lat	Latitude coordinate (extracted) for each unique bay
Long	Longitude coordinate (extracted) for each unique bay
Marker_id	The street marker located next to parking bay with a unique id. This is the key to connect to the sensor_df
BayId	A system generated tag id for each object

Api calls result in all datatypes being imported as strings. Data type changes are made within notebooks. Aggregates are also calculated within notebooks. One key calculation is the occupancy rate of car park bays. An example of the calculation is the sum of DurationSeconds over all parking instances in a time period / total number seconds for that time period in question (e.g. day/month/year). Annual occupancy rate is calculated per individual marker_id in the Southbank Occupancy notebook.

The data source website states that the parking bay location data (including latitude and longitude) and the sensor data can join on the marker_id/streetmarker attribute, which is a unique identifier for an on-street parking bay. This is useful because it allows clear visualisation of parking bays on a map. The bay location data needs cleaning steps to remove nulls and a small number of duplicate marker ids.

Some known data issues exist for the 2017 sensor dataset. A small number of parking instances are excluded at the end of 2017 for parking events that go into 2018. Also, the DurationSeconds columns contains some negative numbers that will need to be removed.

Approach and metrics

Some proposed approaches to the two problem statements are provided.

To remove car parks: Use occupancy rate threshold to flag bays for investigation

Calculate the occupancy rate for each on-street car park bay. Car parks with low occupancy rates (i.e. below a certain threshold) can be flagged for investigation and potential removal. Car parks with low occupancy rates are not being fully utilized and are good candidates for repurposing. Due to the size of the dataset it is not possible to do this across all city regions at once (without access to big-data cloud tools). I have demonstrated an approach and analysis for the Southbank area (initial dataset of ~2.7million rows). Based on the occupancy rates for bays in the Southbank area I have used a threshold of 0.7. This threshold would change for different areas of the city. This can be seen in the associated Tableau file (see Southbank Occupancy files). Visualisations are also provided in the next section.

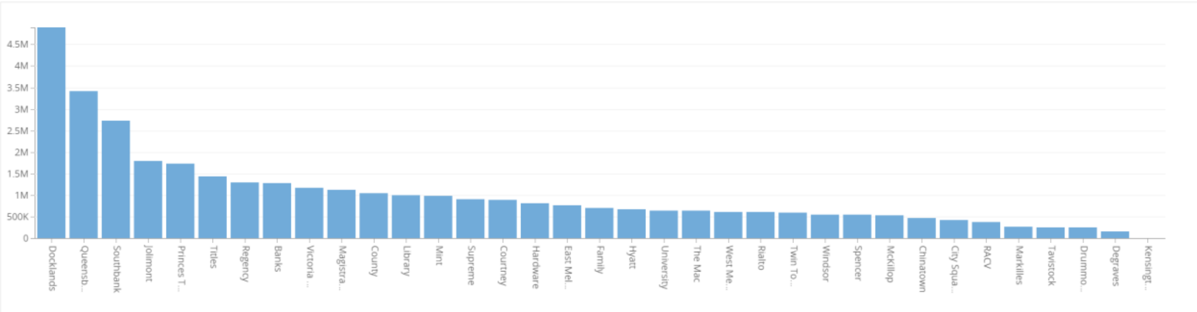
This approach can be easily replicated across each area to find low occupancy car parks for potential removal. This could be useful if a particular area has already been flagged for higher priority car park removal or greening projects.

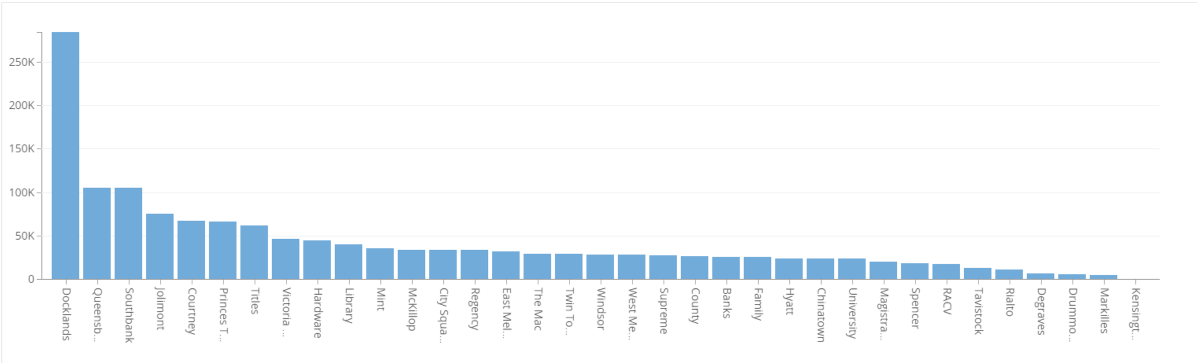
To minimize breaches: Use total counts of violations threshold to flag bays for investigation

It is possible to filter the data for in_violation=True, meaning that parking instances resulting in a violation can be imported. The resulting dataset is manageable in python, but still too large to be useful in Tableau. To help with file size, but still get a view across the city, I have utilized aggregates to group the data. In this case I have aggregated the yearly total count of violations for each parking bay across the city (see All Violations files). From this dataset, a maximum threshold of 2000 violations is proposed. Car parks that show yearly violation counts above this threshold can be flagged for further investigation. Investigation may include an analysis of the park bay location, its proximity to key buildings or landmarks, and the bay restriction for example. Options after investigation may be to change the park bay restriction, increase signage or penalties or further options to be proposed by decision makers.

This approach works well to flag bays for investigation but doesn't allow for drill down analysis of violations over time (because using yearly aggregate result). I have showed an analysis using all parking violation instances over the year for three areas (they must be done separately due to file size limitations). I will detail the Southbank findings here, but I have also analysed Docklands and Victoria Market areas (see respective files). Docklands has the most parking instances and parking violations across the year. Southbank comes in third, but I selected this area because three of the bays in this area were in the top 10 violation counts shown in the aggregate analysis. Victoria Market is in the middle of the pack. Comparing the three allows some trend comparisons across areas over 2017.

The open data page has a function for visualisation. I have showed the total parking instances and total violations per area. The number of violations generally compares well to the number of parking instances. An appreciation of the row sizes of the relative datasets can be seen here. Total violations for Docklands is over 250,000 rows, and total parking instances > 4.5 million rows. This exploratory analysis tool can be used to help decide which areas to focus the analysis on in the initial stages.





Patterns and insights

Southbank Occupancy

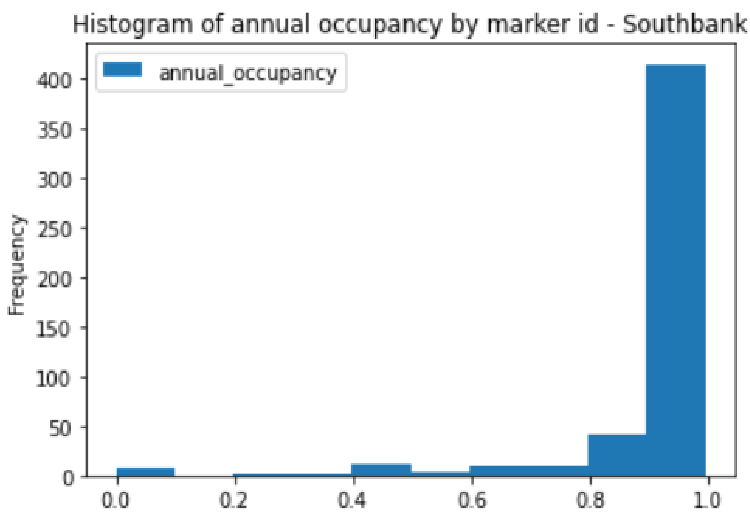
Importing all parking instances into python and selecting key columns gives the output below. A cleaning step present here was to remove all negative values for durationseconds. This is a known issue with the data set. For mathematical operations, the durationseconds is converted to an integer.

	arrivaltime	departuretime	durationseconds	streetmarker	in_violation
0	2017-07-13T20:20:05.000	2017-07-13T22:46:33.000	8788	8668E	False
1	2017-04-21T13:12:01.000	2017-04-21T15:42:32.000	9031	8340E	False
2	2017-01-13T11:52:18.000	2017-01-13T12:02:09.000	591	9427S	False
3	2017-09-16T22:15:44.000	2017-09-16T22:21:06.000	322	9939S	False
4	2017-10-31T12:28:27.000	2017-10-31T12:29:04.000	37	10051S	False

After grouping by streetmarker (unique parking bay) and calculating the occupancy rate for each bay, an example output is shown. As can be seen, three values have occupancy rates above 1.00 which is not possible. Investigation into these three data points showed they had anomalies and they were removed (e.g. multiple very long parking instances on the same day suggesting a sensor or data collection error). It can also be seen that the max duration of these street markers was significantly higher. The same could be done for very low occupancy outliers although it is hard to remove this data without having more information.

streetmarker	max_duration	min_duration	total_duration	annual_occupancy
9138N	21145312	0	3.227753e+10	1023.513732
8000E	28999905	0	5.877521e+09	186.374961
8015W	4308718	0	3.342147e+07	1.059788
9824N	39600	0	3.141170e+07	0.996058
9830N	39600	6	3.136737e+07	0.994653
10034N	62044	0	3.133519e+07	0.993632
9826N	37788	3	3.132758e+07	0.993391
8346E	86136	0	3.132433e+07	0.993288
9956N	39600	4	3.131894e+07	0.993117
9828N	39600	2	3.130739e+07	0.992751

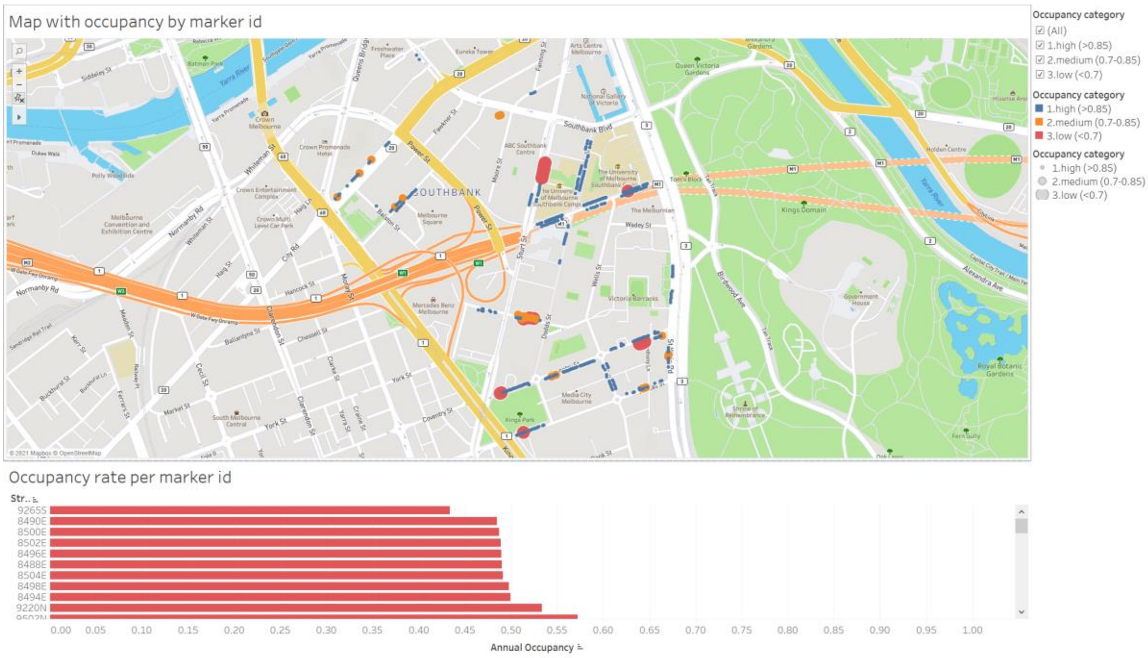
Below is a histogram showing the frequency and distribution of the annual occupancy rates for each marker id. The descriptive statistics are also included for this annual_occupancy calculation. The mean is 0.91 which is higher than I would have expected. The histogram and the interquartile range could be used as statistical tools to help decided a threshold level. I have left that judgement until the visualisation step.



	annual_occupancy
count	499.000000
mean	0.909272
std	0.150238
min	0.000016
25%	0.915477
50%	0.955710
75%	0.978430
max	0.996058

The Tableau interactive dashboard visualizing the park bays by occupancy rate. Upon visualisation in Table I created three grouped to easily see the data. Low occupancy is <0.7, medium 0.7-0.85 and high >0.85. These values can be edited if decision makers want to change these groupings.

The largest red dots show the low occupancy bays. There is a strip of bays along Sturt street, outside the university that have ~0.5 occupancy rates. There are also sections on Coventry and Miles streets that are flagged. These bays are recommended for further investigation and potential removal or repurposing. The lowest occupancy marker ids are shown in the bar chart below the map.



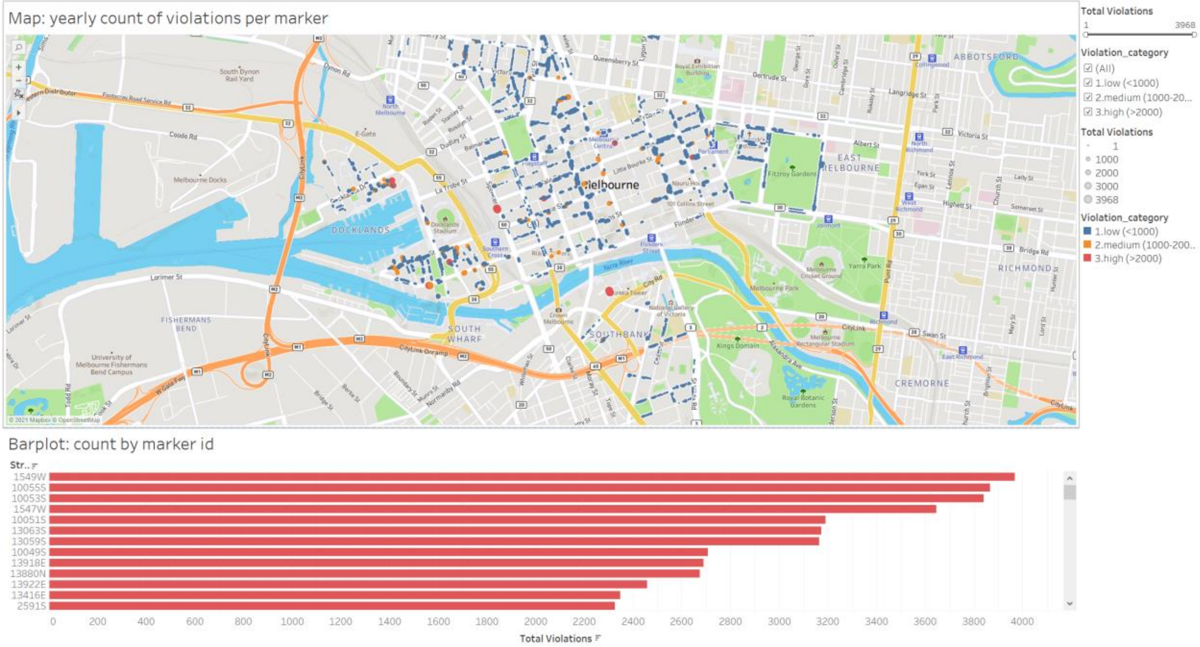
This approach can be applied to other areas across the city.

All violations (aggregate)

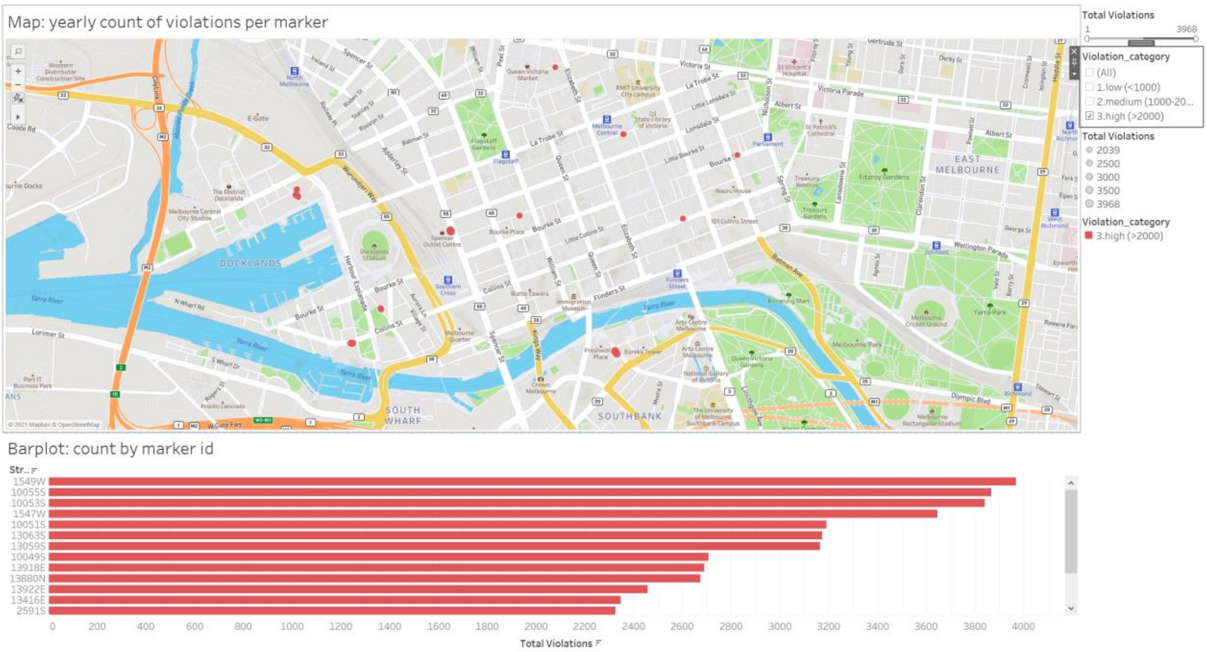
An example output of the grouped and joined dataset for this step is shown below

	streetmarker	total_violations	long	lat
0	10009W	162	144.947719	-37.804563
1	1002E	112	144.962803	-37.818161
2	10041S	623	144.965739	-37.823369
3	10043S	475	144.965666	-37.823379
4	10049S	2707	144.963199	-37.822029

A screenshot of the interactive Tableau dashboard for this analysis is below. Based on the visualisation, categories have been created for low medium and high total violation counts as <1000, 1000-2000, <2000 respectively. The interactive dashboard allows these to be selected to see where the top violations are on the map. These are also shown in the bar chart beneath the map. I have included a second screenshot with only the high violation parking bays selected. These car parks can be selected for further investigation. A recommendation is given on an approach to optimize parking bay restriction in the final section of the report.



Only high violation park bays shown below (flagged bays for investigation).



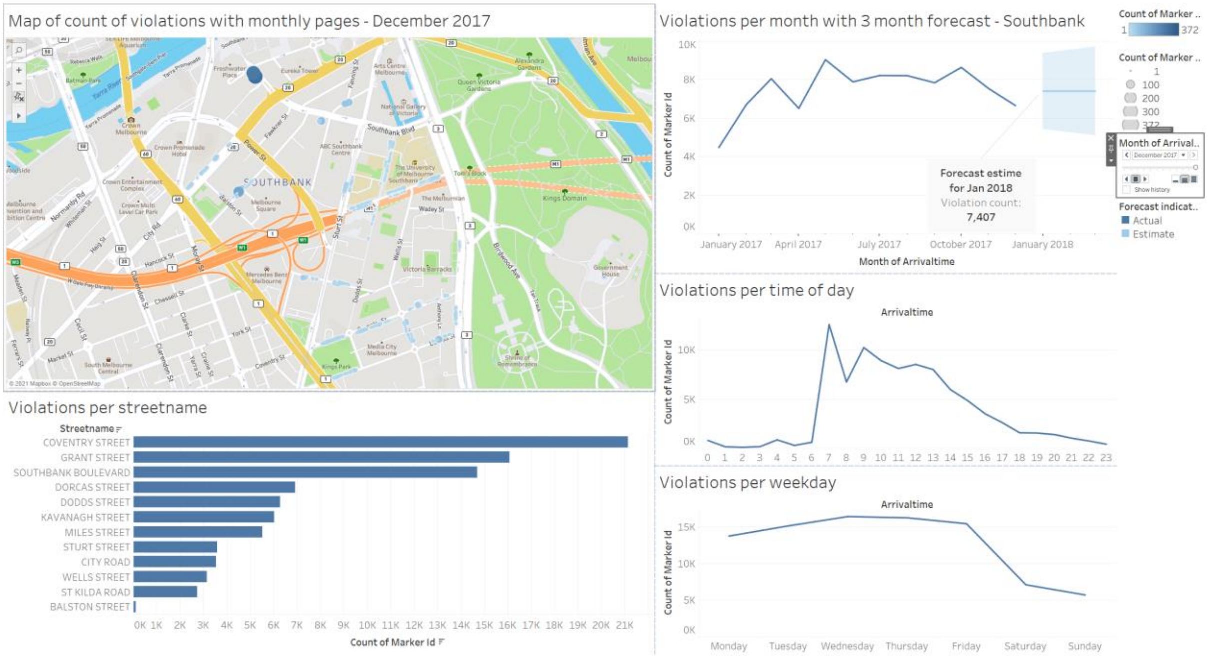
Seeing three of the top 10 bays in Southbank (the three red dots south of the river), I chose to investigate this area further. Docklands and Victoria Market were also analysed but are not included in this report. The analysis files can be accessed for data collection to data visualisation steps.

<u>Southbank Violations (drill down)</u>

All violations for the Southbank region were included in this analysis. An example output of the merged dataset is below. As can be seen, multiple parking instances for the same marker id / deviceid / bay id exist. In Tableau, grouping by the count of the marker_id gives result that can be used as the number of violations in that bay. The benefit of this approach is the time drill down.

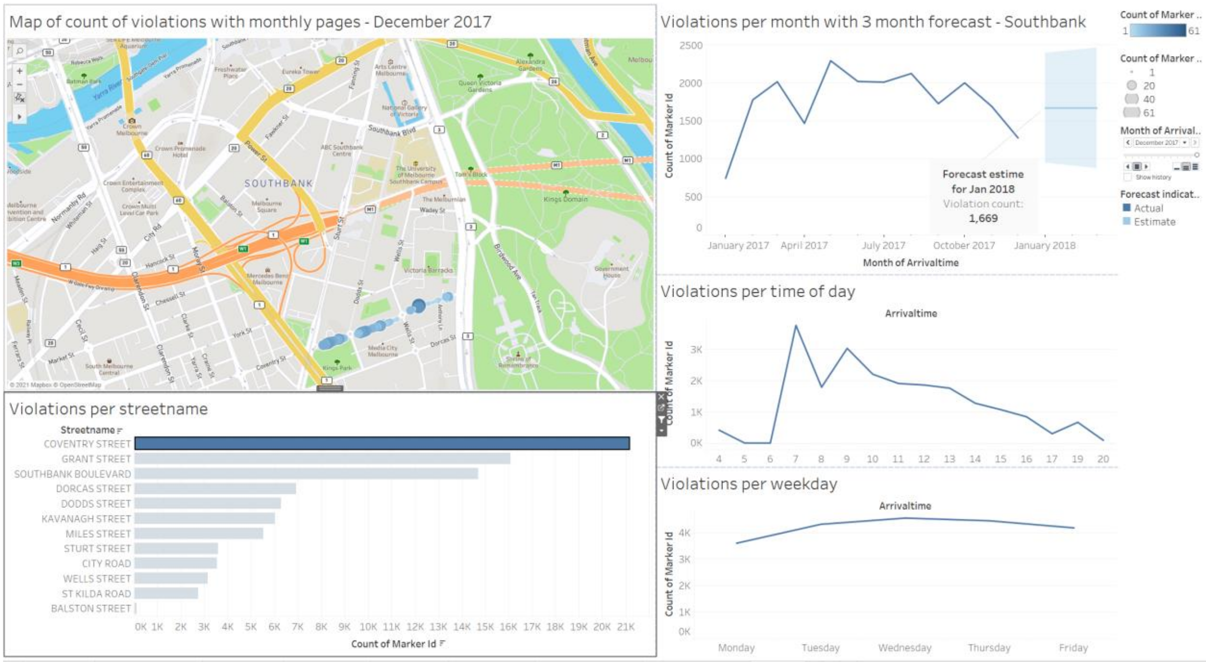
	deviceid	arrivaltime	departuretime	durationseconds	marker_id	area	streetname	bay_id	long	lat
0	21668	2017-11-13T13:08:02.000	2017-11-13T16:14:31.000	11189	8642E	Southbank	KAVANAGH STREET	22717	144.96262	-37.825585
1	21668	2017-06-14T10:15:43.000	2017-06-14T13:53:04.000	13041	8642E	Southbank	KAVANAGH STREET	22717	144.96262	-37.825585
2	21668	2017-10-10T07:30:00.000	2017-10-10T18:30:00.000	39600	8642E	Southbank	KAVANAGH STREET	22717	144.96262	-37.825585
3	21668	2017-02-13T14:38:11.000	2017-02-13T18:30:00.000	13909	8642E	Southbank	KAVANAGH STREET	22717	144.96262	-37.825585
4	21668	2017-07-31T07:30:00.000	2017-07-31T10:52:35.000	12155	8642E	Southbank	KAVANAGH STREET	22717	144.96262	-37.825585

The below dashboard utilises Tableau functionality to include new plots.

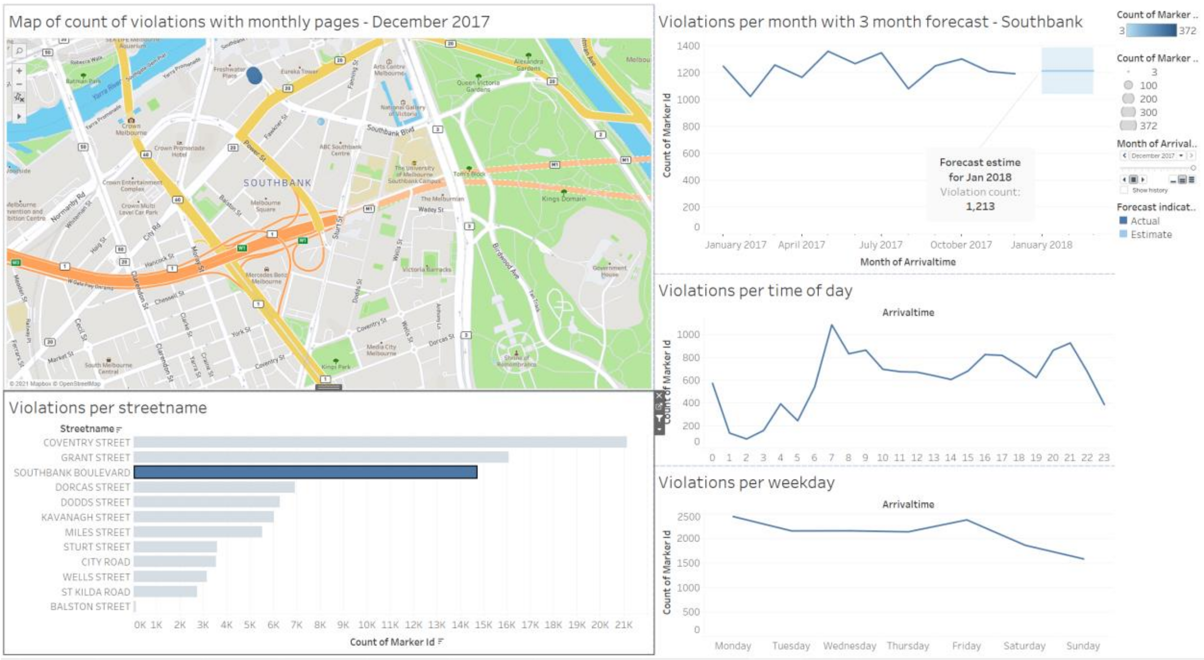


A monthly page function shows the change in violations for each bay over 2017. The monthly trends with estimates are provided (to be further analysed in the next section). The hourly plot shows that the most violations occurred at 7am with a slow decrease over the day. As expected there are less violations from 6pm-6am were there are less restrictions. This same trend is seen in the weekday chart where weekends see less violations.

The three high violation parking bays sit on Southbank Boulevard (flagged for investigation with >2000). This dashboard makes use of a street filter. Clicking on the street updates the other plots. Notedly, Coventry street has higher total violations than Southbank Boulevard. On inspection this makes sense as more car parks are available on that street. It also suggests that just total violations may not be the only thing to consider for car park removal. Using the monthly page shows visual changes over the months of 2017 for that street. The monthly and hourly trends for Coventry street are similar to the total violation trends, but there are no weekend violations, likely due to no weekend restrictions.



Below is Southbank Boulevard which shows different time trends. Violations are seen in the evening and on the weekend, likely due to different restrictions on parts of that street. Violation behaviour appears to be different between the two streets.

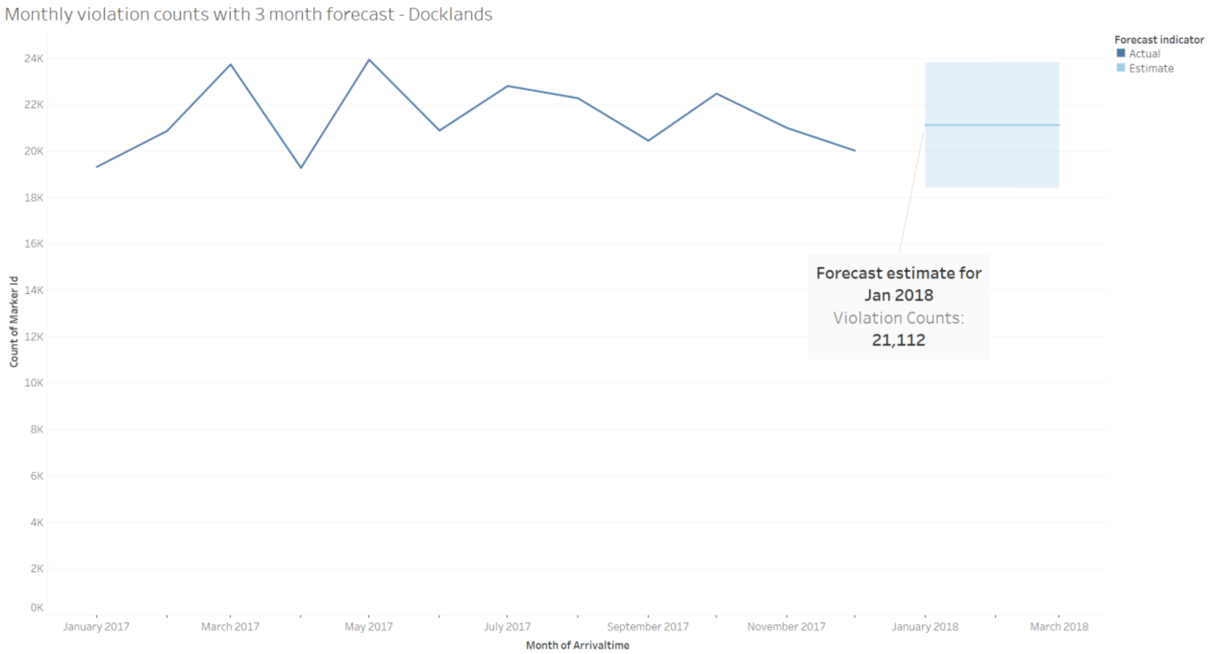


This shows that different approaches can be taken to flag car parks or streets for further investigation. For example, even though Coventry street didn't have individual bays with high violations, but the street contributes more to violations than Southbank Boulevard. This second approach also allows analysis of key streets within areas across the city.

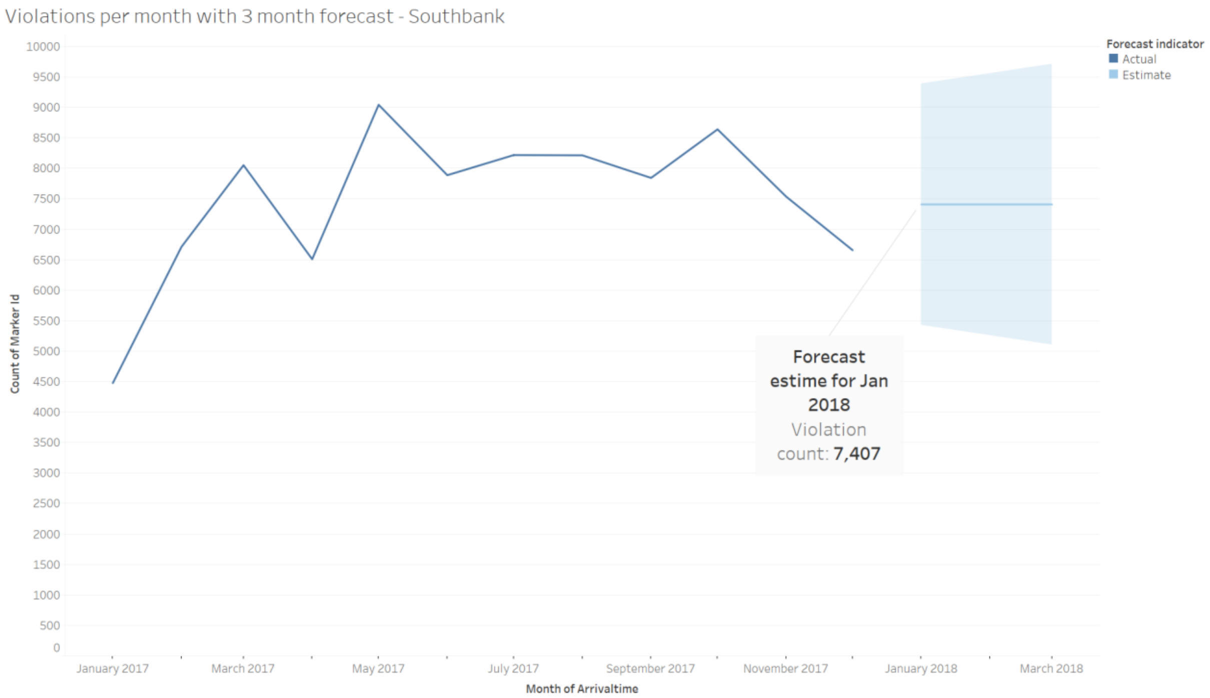
Time trends and predictions

The following three charts show total monthly violation counts for 2017 for Docklands, Southbank, and Victoria Market areas. The all show a 3-month forecast with an estimate value for January 2018.

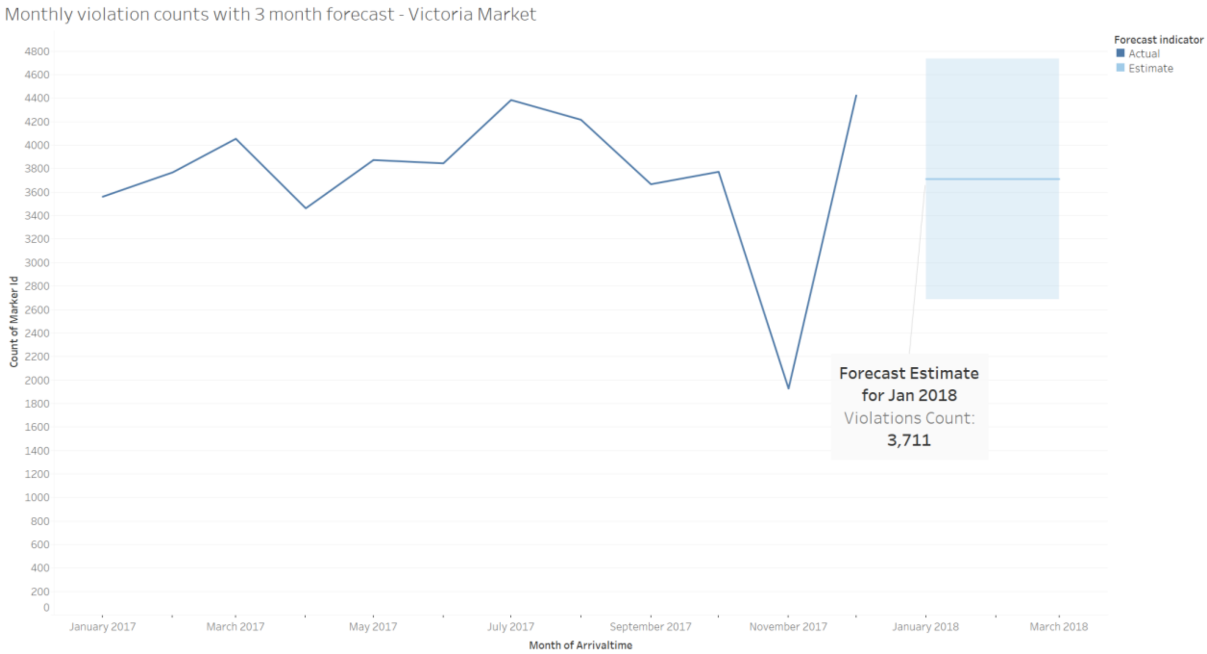
Docklands shows a relatively stable violation counts with most of the year having between 20,000 and 24,000 violations each month. Southbank has a yearly average of around 7,500 violations per month and Victoria around 3800 in comparison.



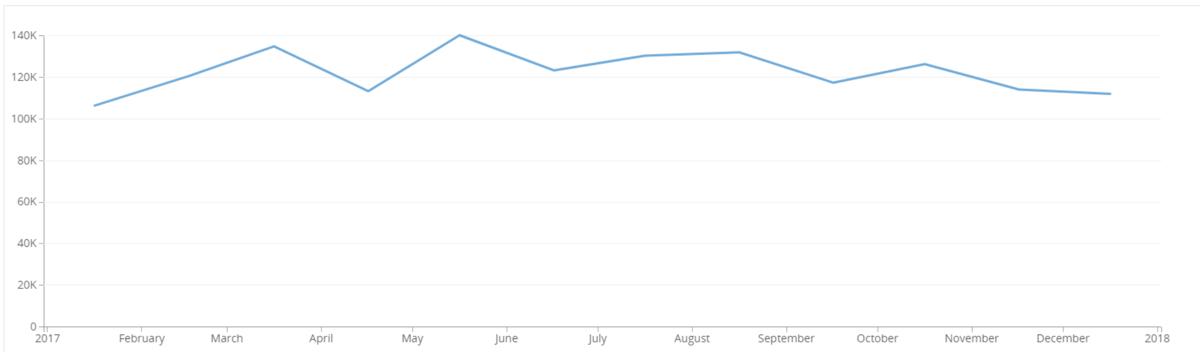
Southbank has some drop offs at the beginning and the end of the year in comparison with Docklands. This might suggest a cyclical trend in which case the forecast analysis may not be useful. From this trend it would be expected that January would have a lower number than then given estimate of ~7400.



Victoria Market shows a similar steady trend to Docklands until November where there is a large drop. This is unusual and should be investigated. A further plot is shown below which is the trend of all violations taken from the source website. It shows no drop off in November.



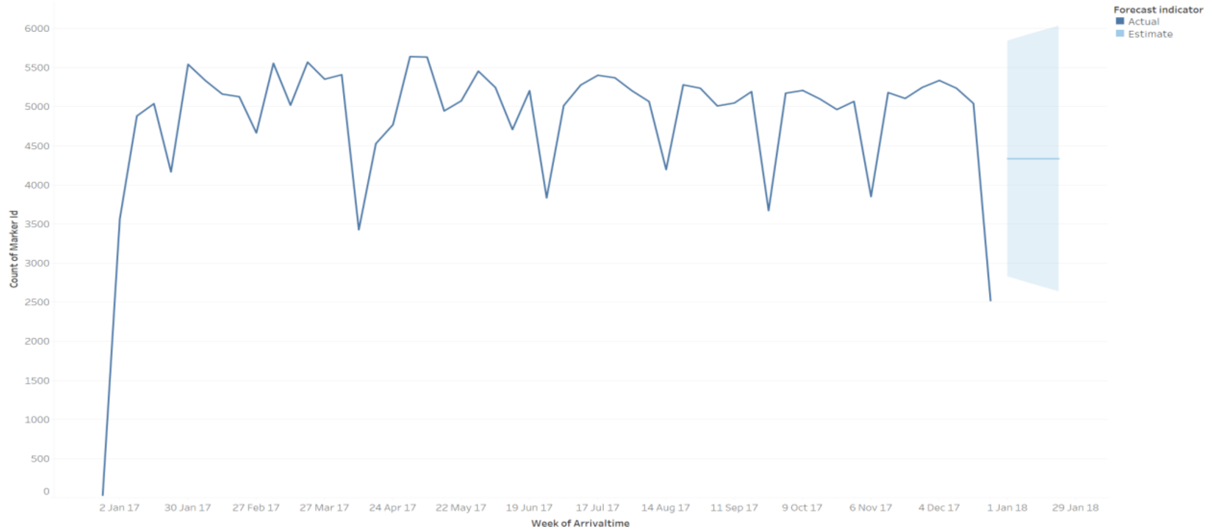
The plot of all violations below does show some evidence of a cyclical trend with slight drop off at the beginning and end of the year. There is also a drop off in April that coincides with Easter. Easter weekend was the 14-17 April in 2017 (Office Holidays, 2021). With some suggestion of cyclical or seasonal trends, it would be better to include more data (from 2017 to date) and include seasonal adjustments for predictions or forecasting of violations.



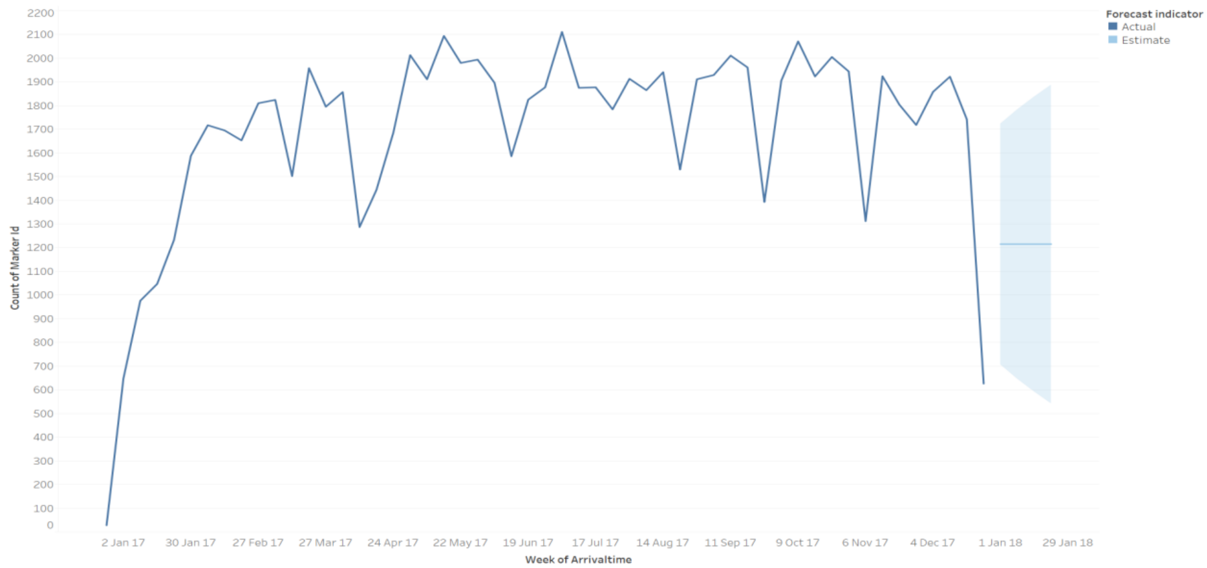
The following three plots show weekly violations for 2017 with 4-week forecasts for Docklands, Southbank, and Victoria Market areas. Docklands and Southbank show drop offs at the beginning and end of the year. Interestingly Southbank shows a slower uptake of violation counts at the beginning of the year. The Victoria Market weekly trend highlights the November drop more clearly. An investigation into data collection for two weeks around the 13 November could help uncovered any issues. This period may also have a reasonable explanation for lower violations so the data should not be excluded until an investigation is done.

Overall, forecasts may give an estimate for early 2018, but more yearly data should be included to increase the usefulness of these predictions.

Weekly violation counts with 4 week forecast - Docklands



Violations per week with 4 week forecast - Southbank

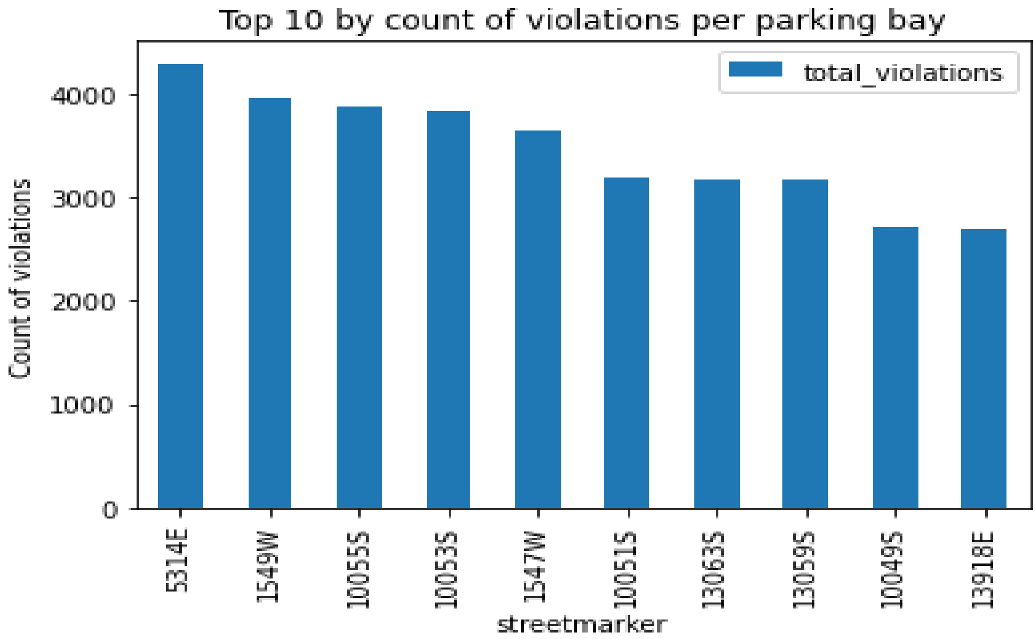


Weekly violation counts with 4 week forecast - Victoria Market

Drawbacks of the analysis

File size and computer processing has limited some of the analysis in this project. Having the ability to work on the full 2017 dataset may have proved fruitful. This may be possible with big-data cloud tools. In the time frame given it was only possible to analyse one area for occupancy rates, but this would have been useful to do across the whole city. Filtering Api calls does work for limiting data, but it often meant it was hard to get an overall view of the city. Using aggregates was one approach, but that can limit the drill down functionality of visualisation tools.

As with all analyses there is often data loss through cleaning and merging. This was also the case with this project. Known data issues and missing values meant removal of some data. The most significant impact for this analysis was joining the location dataset to the 2017 sensor parking instances dataset. The unique streetmarker key had some mismatches and therefore excluded some data points. The analysis and conclusions drawn should be seen in this context. An example of a drawback of the approach is highlighted below.



This plot shows the top 10 aggregate violation counts on the whole 2017 dataset. The parking bay with the most violations was 5314E (4299 violations) located on William Street. It does not have a matching street marker in the locations data set (to get lat/long points for map visualisation). This means it is excluded when joining and not included in the Tableau visualisation. For each merge in the python notebooks, marker id's that won't be included after the join are noted (individual investigation is possible). Perhaps there is a reason the parking bay noted above has no street marker. Checking this marker id in the dataset showed it has a P10mins AOT restriction. Being a 10min parking spot could mean it is not included as an on-street parking bay. Perhaps street markers could be included or updated if they already exist so that this analysis is more complete. This car park could be added to the other flagged bays for investigation. Further, an analysis into the relationship between bay restrictions and violation counts could be done as further work.

Recommendations and further work

Recommendations and further work are given here. The first suggestion is to extend the occupancy rate analysis across all areas in the city. Comparing thresholds across areas could help to see if some areas have more low occupancy bays than others. Areas with a lot of low occupancy on-street parking bays could be investigated for improvement projects such as greening or introducing cycle lanes. A whole city approach could also be taken with the use of big-data cloud tools. This may provide a city wide metric similar to the violation counts.

Investigating occupancy rates over time (hour of day, weekday, weeks of year, seasonal) could show useful trends like in the violation work.

A recommendation to optimize bay restrictions is to use the flagged violation bays (>2000) in an A/B test experiment. Increasing the restriction time on some bays and decreasing on others may show a reduction or increase in violations over time. This could also be compared to historical data taking into account seasonal trends.

An investigation into the variables that contributing to occupancy rate and violation counts would be interesting. This may require the inclusion of new data. Using key variables, a classification model could be used to rank bay across the city by importance, effectiveness or to flag them as potential problem bays for removal or bay restriction optimisation.

Finally, a larger scale project may be to investigate dynamic car park pricing. This may be possible by utilizing the live sensor parking data, occupancy rates and a demand-response pricing strategy. San Francisco already have a program in place for meters and garage prices to match demand. Parking prices are higher when occupancy rates are higher, and prices are lower during low occupancy periods (SFMTA, 2021). This would require a business case as methodology and infrastructure would need to be built for applying dynamic pricing.

References

City of Melbourne, 2021, Transport Strategy 2030, https://www.melbourne.vic.gov.au/parking-and-transport/transport-planning-projects/pages/transport-strategy.aspx

City of Melbourne, 2021, On-street Parking Sensor Data – 2017, https://data.melbourne.vic.gov.au/Transport/On-street-Car-Parking-Sensor-Data-2017/u9sa-j86i

City of Melbourne, 2021, On-street Parking Bay Sensors, accessing data, https://dev.socrata.com/foundry/data.melbourne.vic.gov.au/vh2v-4nfs

City of Melbourne, 2021, On-street Parking Bay spatial polygons, https://data.melbourne.vic.gov.au/Transport/On-street-Parking-Bays/crvt-b4kt

City of Melbourne, 2021, On-street Parking Bay Restrictions, https://data.melbourne.vic.gov.au/Transport/On-street-Car-Park-Bay-Restrictions/ntht-5rk7

Office Holidays, 2021, https://www.officeholidays.com/countries/australia/2017

City and County of San Francisco, 2021, https://www.sfmta.com/demand-responsive-parking-pricing

Appendix

Summary of submitted files for reference

Group	File name	Description
1. Southbank Occupancy	Southbank Occupacy.ipynb	All parking instances for Southbank. Aggregate by marker_id, use aggregates to calculate occupancy rate
	Southbankoccupancy.csv	Export csv
	SouthbankOccupancy.twb	Visualise occupancy rates for each parking bay in Southbank. Can use to see low occupancy bays for investigation for removal
2. All violations	All_violations.ipynb	All violations for all areas. Aggregate total counts over the year for each parking bay
	yearly_violations_agg_all_bays.csv	Export file from python to use in Tableau
	All_violations.twb	Visualisations including dashboard with calculated field category. Can use to show high violation bays for investigation
3. Southbank violations	Southbank Violations.ipynb	Full year violations for area Southbank
	SouthbankViolations.csv	Export file from python to Tableau
	SouthbankViolations.twb	Visualisation of Southbank violations. Includes monthly chart, map with monthly pages, forecast prediction based on 2017 data.
4. Docklands violations	Docklands Violations.ipynb	Full year violations for area Docklands
	DocklandsViolationswithlocation.csv	Export csv
	Docklands_all_violations.twb	Visualisation of Docklands violations. Includes monthly chart, map with monthly pages, forecast prediction based on 2017 data.
5. Victoria Market violations	Victoria Market Violations.ipynb	Full year violations for area Victoria Market
	VictoriaMarketViolations.csv	Export csv
	VictoriaMarketViolations..twb	Visualisation of Victoria Market violations. Includes monthly chart, map with monthly pages, forecast prediction based on 2017 data.