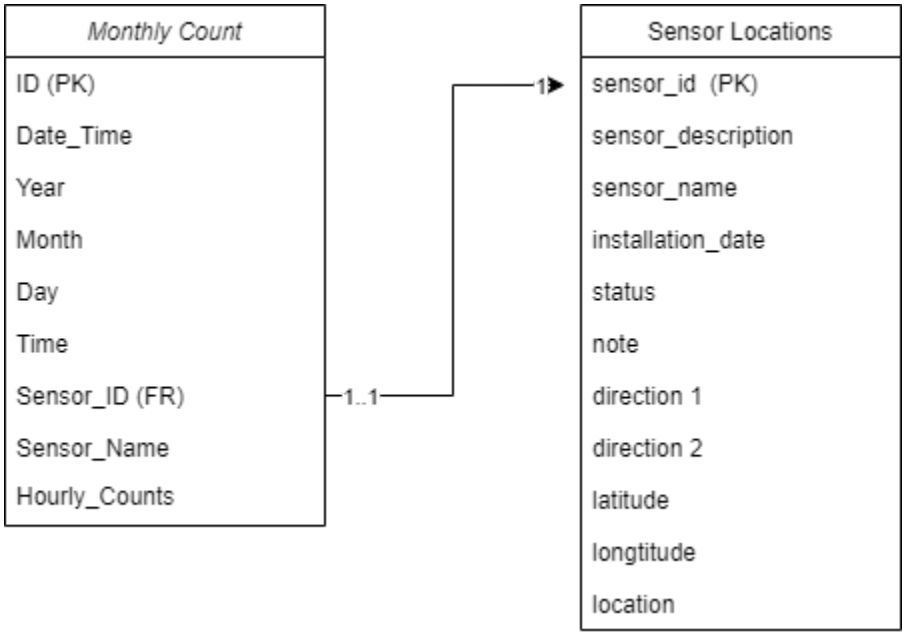
Pedestrian count coding assignment

This is the documentation of a coding task. This document contains the data modeling, data loading type, approach, high-level architecture, data issue, statistic output, and future prospects.

Data modeling Diagram



Before we proceed to how I approached this problem, let's discuss the data modeling and the data load type choice that I made.

This is the data model that I have derived from the dataset. The monthly count dataset contains an hourly count for each sensor in Melbourne. Each row contains ID as a primary key and Sensor ID as a foreign key. It has multiple date-time columns such as Date_TIme, Year, Month, Day, and Time. This is helpful for developers to further manipulate the data and perform analysis better.

In the sensor location data model, it has all the sensor detailed information such as where is it placed, when it’s installed and the status of the sensor. This table has sensor_id as a primary key which can be linked to the Monthly count table.

They have a one-to-one relationship because each row in the hourly count table can have only one sensor id and each sensor only has one name/location.

Data Loading Type

Monthly Count table

Column	Load-Type
ID	Integer
Date_Time	Text
Year	Integer
Month	Text
Mday	Integer
Day	Integer
Time	Text
Sensor_ID	Integer
Sensor_Name	Text
Hourly_Counts	Integer

Sensor Location table

Column	Load-Type
sensor_id	Integer
sensor_description	Text
sensor_name	Text
installation_date	Datetime
status	Text
note	Text

direction_1	Text
direction_2	Text
latitude	Float
longitude	Float
Location	Text

Approach

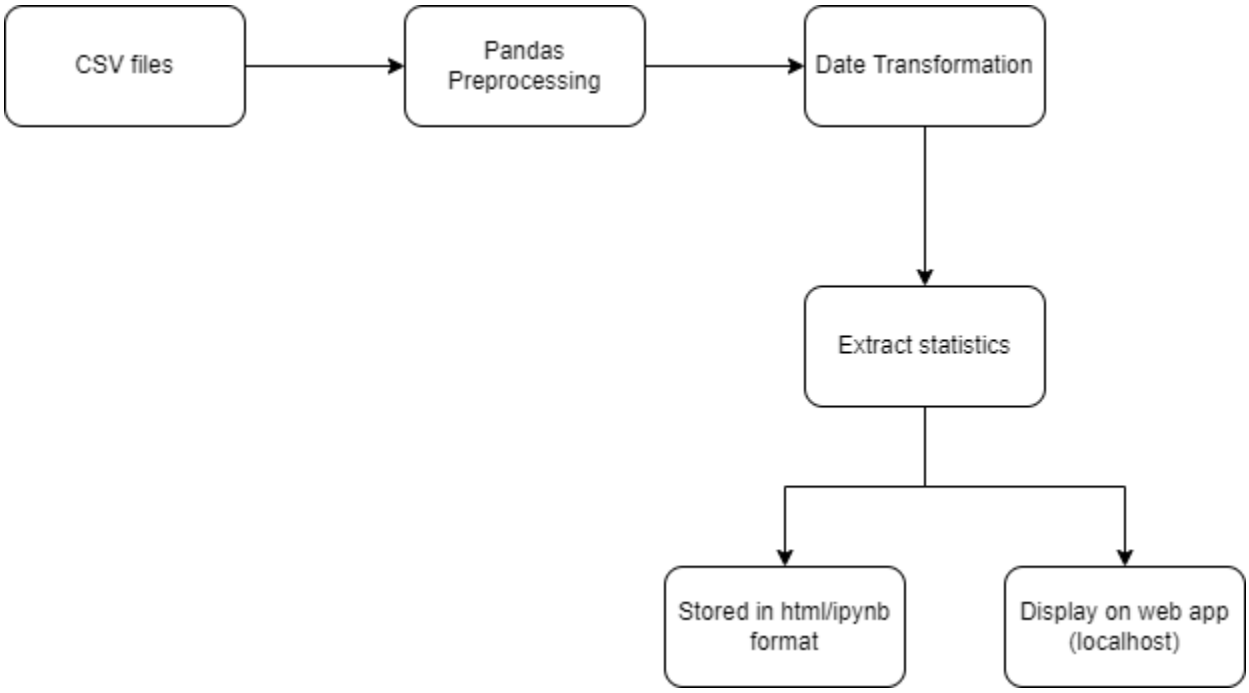
For this problem, I used pandas and dash Plotly as my primary components to get the stats questions asked and performed analysis. I have created both a dash web app and a python notebook for this problem.

For the preprocessing step, I filtered out the inactive or removed sensors from the hourly count which are 11 sensors removed and 2 sensors inactivated. Since the note column in the sensor location table is mostly NaN. It means there are no notes of when it's removed and why. I made the decision to not included 13 sensors that are inactive/Removed since it has too much uncertainty and would affect the stats extraction later on.

For the statistics extraction, I used pandas to do preprocessing and do a group-by for each question since that is the easiest way to get the information that we want.

Lastly, for visualization purposes, I created a dash Plotly web app for interactive viewing of statistics extracted from these datasets. The reason is dash web app required less time to implement and it's also in Python so that simplified the process.

Architecture



Data Issue

As I mentioned, there are some sensors that are either not activated or removed. The problem with these rows is that the note column is not very informative as it's not telling us when it's got inactivated or removed.

Statistics Extraction

Top 10 (most pedestrians) locations by day

To extract this information, I use a group by in Pandas and group them by day.

	Day	Sensor_ID	Sensor_Name	Hourly_Counts
82	Monday	4	Town Hall (West)	19272251
84	Monday	6	Flinders Street Station Underpass	17752947
81	Monday	3	Melbourne Central	15577022
80	Monday	2	Bourke Street Mall (South)	14689966
83	Monday	5	Princes Bridge	14521729
79	Monday	1	Bourke Street Mall (North)	14218144
97	Monday	24	Spencer St-Collins St (North)	11777417
87	Monday	9	Southern Cross Station	9310691
104	Monday	35	Southbank	8602186
100	Monday	28	The Arts Centre	8334224

	Day	Sensor_ID	Sensor_Name	Hourly_Counts
398	Tuesday	4	Town Hall (West)	19457168
400	Tuesday	6	Flinders Street Station Underpass	18461430
397	Tuesday	3	Melbourne Central	16073224
396	Tuesday	2	Bourke Street Mall (South)	14429616
399	Tuesday	5	Princes Bridge	14396109
395	Tuesday	1	Bourke Street Mall (North)	14273069
413	Tuesday	24	Spencer St-Collins St (North)	12543187
403	Tuesday	9	Southern Cross Station	10230535
420	Tuesday	35	Southbank	8956773
416	Tuesday	28	The Arts Centre	8456510

	Day	Sensor_ID	Sensor_Name	Hourly_Counts
477	Wednesday	4	Town Hall (West)	20360823
479	Wednesday	6	Flinders Street Station Underpass	19002309
476	Wednesday	3	Melbourne Central	16657083
478	Wednesday	5	Princes Bridge	15592856
475	Wednesday	2	Bourke Street Mall (South)	15343311
474	Wednesday	1	Bourke Street Mall (North)	15043324
492	Wednesday	24	Spencer St-Collins St (North)	12922557
482	Wednesday	9	Southern Cross Station	10305852
495	Wednesday	28	The Arts Centre	9184199
499	Wednesday	35	Southbank	9181249

	Day	Sensor_ID	Sensor_Name	Hourly_Counts
319	Thursday	4	Town Hall (West)	21187600
321	Thursday	6	Flinders Street Station Underpass	19413550
318	Thursday	3	Melbourne Central	17019854
317	Thursday	2	Bourke Street Mall (South)	16322770
316	Thursday	1	Bourke Street Mall (North)	15965870
320	Thursday	5	Princes Bridge	15704870
334	Thursday	24	Spencer St-Collins St (North)	12928593
324	Thursday	9	Southern Cross Station	10424560
341	Thursday	35	Southbank	9364584
337	Thursday	28	The Arts Centre	9160781

	Day	Sensor_ID	Sensor_Name	Hourly_Counts
3	Friday	4	Town Hall (West)	24393876
5	Friday	6	Flinders Street Station Underpass	21211433
2	Friday	3	Melbourne Central	19853833
1	Friday	2	Bourke Street Mall (South)	18800045
4	Friday	5	Princes Bridge	18367835
0	Friday	1	Bourke Street Mall (North)	18348011
18	Friday	24	Spencer St-Collins St (North)	12951080
25	Friday	35	Southbank	10917020
21	Friday	28	The Arts Centre	10007238
8	Friday	9	Southern Cross Station	9761409

	Day	Sensor_ID	Sensor_Name	Hourly_Counts
161	Saturday	4	Town Hall (West)	22499236
160	Saturday	3	Melbourne Central	19012837
162	Saturday	5	Princes Bridge	18233579
159	Saturday	2	Bourke Street Mall (South)	17003142
158	Saturday	1	Bourke Street Mall (North)	16748863
163	Saturday	6	Flinders Street Station Underpass	15314200
183	Saturday	35	Southbank	10878637
179	Saturday	28	The Arts Centre	9753895
188	Saturday	41	Flinders La-Swanston St (West)	7017338
176	Saturday	24	Spencer St-Collins St (North)	6912910

	Day	Sensor_ID	Sensor_Name	Hourly_Counts
240	Sunday	4	Town Hall (West)	18172940
239	Sunday	3	Melbourne Central	15544124
241	Sunday	5	Princes Bridge	15480437
237	Sunday	1	Bourke Street Mall (North)	13626411
238	Sunday	2	Bourke Street Mall (South)	13405435
242	Sunday	6	Flinders Street Station Underpass	12931378
262	Sunday	35	Southbank	9037869
258	Sunday	28	The Arts Centre	8853109
253	Sunday	21	Bourke St-Russell St (West)	5789151
267	Sunday	41	Flinders La-Swanston St (West)	5716816

Top 10 (most pedestrians) locations by Month

To extract this information, I use a group by in Pandas and group them by Month.

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
294	January	4	Town Hall (West)	11967639
295	January	5	Princes Bridge	10796073
296	January	6	Flinders Street Station Underpass	10279628
292	January	2	Bourke Street Mall (South)	9106886
293	January	3	Melbourne Central	8433264
291	January	1	Bourke Street Mall (North)	8136291
316	January	35	Southbank	6279974
312	January	28	The Arts Centre	6214671
309	January	24	Spencer St-Collins St (North)	6096117
297	January	7	Birrarung Marr	4494074

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
217	February	4	Town Hall (West)	11156945
219	February	6	Flinders Street Station Underpass	10292516
218	February	5	Princes Bridge	9493960
216	February	3	Melbourne Central	9004266
214	February	1	Bourke Street Mall (North)	8335328
215	February	2	Bourke Street Mall (South)	8134134
232	February	24	Spencer St-Collins St (North)	6591467
235	February	28	The Arts Centre	5657723
239	February	35	Southbank	5462142
222	February	9	Southern Cross Station	4481588

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
512	March	4	Town Hall (West)	12655259
511	March	3	Melbourne Central	11737486
514	March	6	Flinders Street Station Underpass	11419121
513	March	5	Princes Bridge	11001403
509	March	1	Bourke Street Mall (North)	9643824
510	March	2	Bourke Street Mall (South)	9017586
527	March	24	Spencer St-Collins St (North)	7149300
534	March	35	Southbank	5930487
530	March	28	The Arts Centre	5699004

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
3	April	4	Town Hall (West)	12463968
2	April	3	Melbourne Central	10862976
4	April	5	Princes Bridge	10660422
5	April	6	Flinders Street Station Underpass	10135936
0	April	1	Bourke Street Mall (North)	9194871
1	April	2	Bourke Street Mall (South)	8815472
18	April	24	Spencer St-Collins St (North)	6215026
21	April	28	The Arts Centre	6090009
25	April	35	Southbank	6002385
8	April	9	Southern Cross Station	3938379

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
589	May	4	Town Hall (West)	11591495
591	May	6	Flinders Street Station Underpass	10787077
588	May	3	Melbourne Central	10615667
586	May	1	Bourke Street Mall (North)	9774547
590	May	5	Princes Bridge	9561337
587	May	2	Bourke Street Mall (South)	8901477
604	May	24	Spencer St-Collins St (North)	6009446
611	May	35	Southbank	5862252
594	May	9	Southern Cross Station	5034469
607	May	28	The Arts Centre	5004039

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
435	June	4	Town Hall (West)	11934368
434	June	3	Melbourne Central	10287904
437	June	6	Flinders Street Station Underpass	9867379
432	June	1	Bourke Street Mall (North)	9455301
436	June	5	Princes Bridge	8997795
433	June	2	Bourke Street Mall (South)	8818621
450	June	24	Spencer St-Collins St (North)	6139724
457	June	35	Southbank	5242526
453	June	28	The Arts Centre	5039946
440	June	9	Southern Cross Station	4525398

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
370	July	4	Town Hall (West)	12555809
369	July	3	Melbourne Central	10445330
372	July	6	Flinders Street Station Underpass	10203582
367	July	1	Bourke Street Mall (North)	9103779
368	July	2	Bourke Street Mall (South)	9048971
371	July	5	Princes Bridge	8531264
385	July	24	Spencer St-Collins St (North)	6016559
392	July	35	Southbank	4994974
388	July	28	The Arts Centre	4954330
375	July	9	Southern Cross Station	4798905

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
80	August	4	Town Hall (West)	11675587
79	August	3	Melbourne Central	10477135
82	August	6	Flinders Street Station Underpass	9214815
77	August	1	Bourke Street Mall (North)	8299742
78	August	2	Bourke Street Mall (South)	8282631
81	August	5	Princes Bridge	7861189
95	August	24	Spencer St-Collins St (North)	5933432
85	August	9	Southern Cross Station	4647033
98	August	28	The Arts Centre	4493078
102	August	35	Southbank	4477946

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
799	September	4	Town Hall (West)	11475209
798	September	3	Melbourne Central	9902555
801	September	6	Flinders Street Station Underpass	9602620
797	September	2	Bourke Street Mall (South)	8286561
796	September	1	Bourke Street Mall (North)	8192967
800	September	5	Princes Bridge	7823066
814	September	24	Spencer St-Collins St (North)	6126071
817	September	28	The Arts Centre	4742644
821	September	35	Southbank	4661031
804	September	9	Southern Cross Station	4214830

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
733	October	4	Town Hall (West)	11152763
735	October	6	Flinders Street Station Underpass	10333881
730	October	1	Bourke Street Mall (North)	8737215
731	October	2	Bourke Street Mall (South)	8675275
732	October	3	Melbourne Central	8260476
734	October	5	Princes Bridge	7877358
748	October	24	Spencer St-Collins St (North)	6728313
755	October	35	Southbank	5276274
751	October	28	The Arts Centre	5036796
738	October	9	Southern Cross Station	4539958

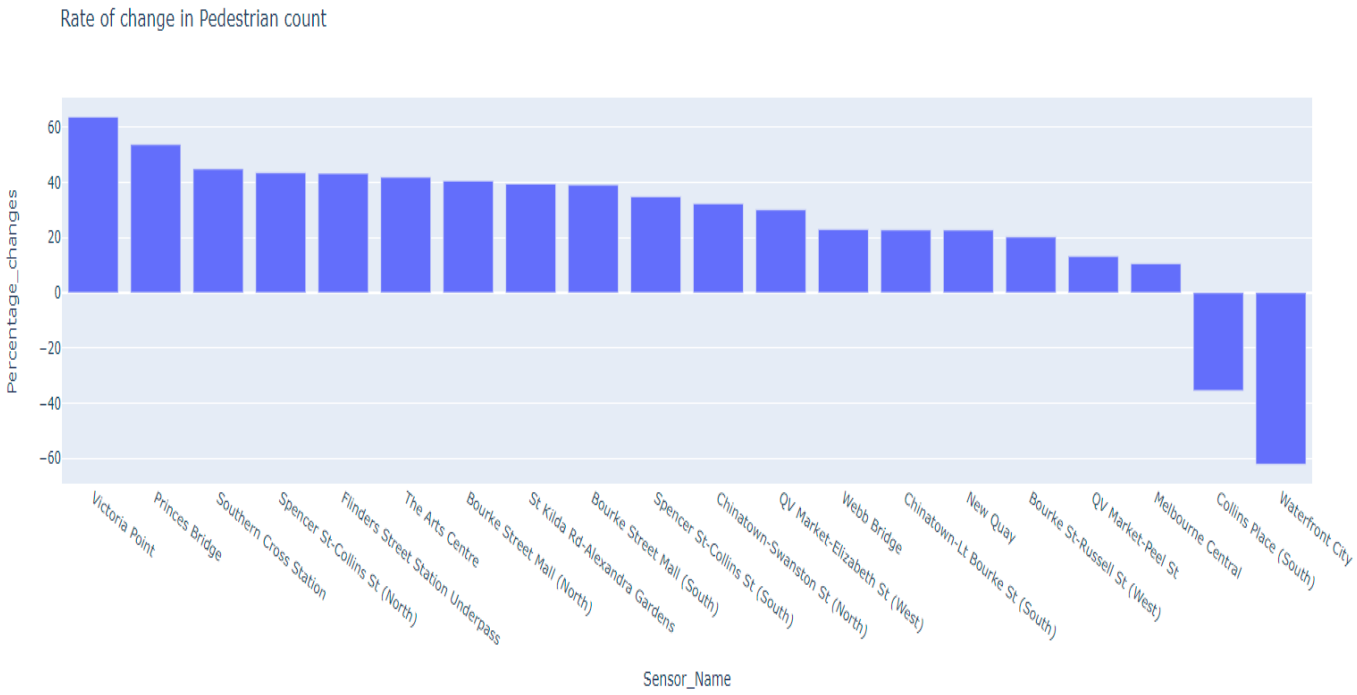
	Month	Sensor_ID	Sensor_Name	Hourly_Counts
667	November	4	Town Hall (West)	12136661
669	November	6	Flinders Street Station Underpass	10598560
665	November	2	Bourke Street Mall (South)	9772438
666	November	3	Melbourne Central	9752026
668	November	5	Princes Bridge	8942629
664	November	1	Bourke Street Mall (North)	8871745
682	November	24	Spencer St-Collins St (North)	6661345
689	November	35	Southbank	5660775
685	November	28	The Arts Centre	5351393
672	November	9	Southern Cross Station	4265450

	Month	Sensor_ID	Sensor_Name	Hourly_Counts
145	December	4	Town Hall (West)	14578191
143	December	2	Bourke Street Mall (South)	13134233
147	December	6	Flinders Street Station Underpass	11352132
146	December	5	Princes Bridge	10750919
142	December	1	Bourke Street Mall (North)	10478082
144	December	3	Melbourne Central	9958892
167	December	35	Southbank	7087552
160	December	24	Spencer St-Collins St (North)	5946613
163	December	28	The Arts Centre	5466323
172	December	41	Flinders La-Swanston St (West)	5116437

Which location has shown most decline due to lockdowns in last 2 years

I filtered this data from 1 January 2020 onward. The reason is that the lockdown started around March 2020 so I want to get some numbers before the lockdown period and compare them after the lockdown. The rate of change is calculated by using the January hourly count minus the latest hourly count in the date range of June 2022 and divided by the January hourly count to get the percentage.

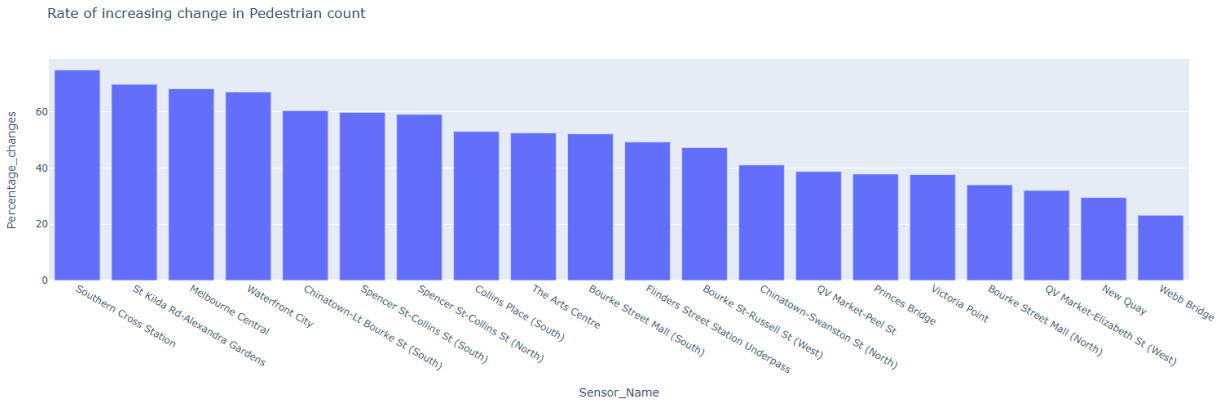
The most declined pedestrian count location is Victoria point with down 63.717622% from the January hourly count.



Which location has most growth in last year

Since the latest data in hourly count is June 2022. I filtered the data from June 2021 to June 2022. Performed a similar calculation as above except that I swapped the variable from oldest data minus the latest data, I used the latest data minus the oldest data instead.

The most growth in pedestrian count location is Southern Cross Station with 74.8% increase from last year.



You can view these graphs in higher definition and see my code in analysis.html for further understanding of my approach.

Future loading and querying

The method that I am using to get the desired statistic is not ideal for incremental data inserted every month. The reason is that I have to reload an entire ipython notebook to generate all the data. I would create a database with staging schema and curated schema. In which the code will read directly from the database and we don’t have to worry about incremental changes.

Further statistics extraction

I think we can create a heatmap of all locations indicated in the hourly count to get a sense of how many pedestrians are concentrated at which point in time. If we also have a tram or train dataset, we can extrapolate how it affects people's count in these sensor locations.