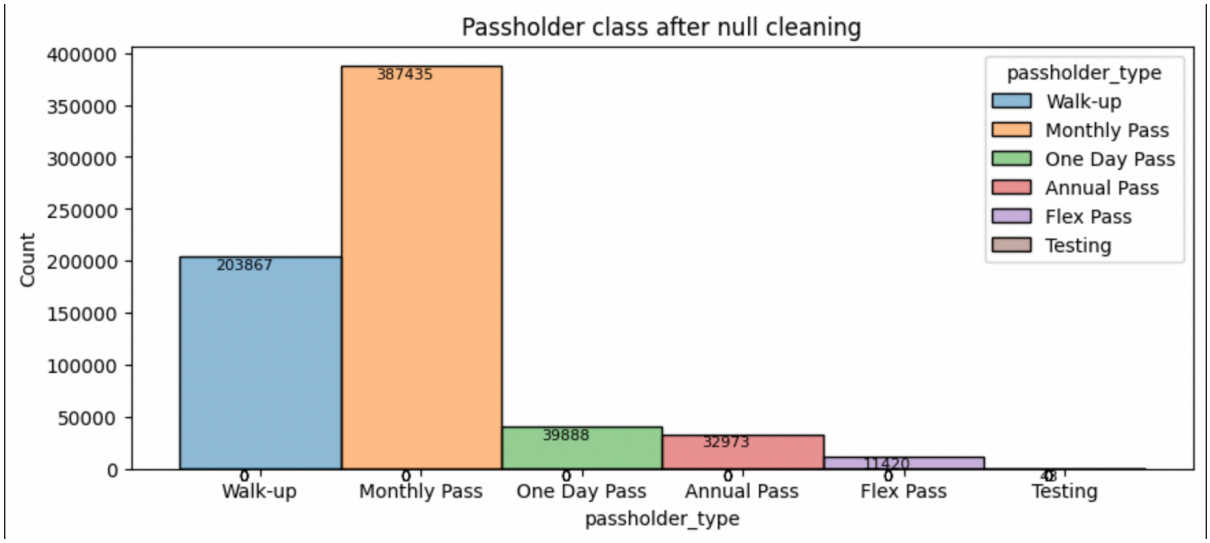
agustin genoud
@linkedin
@github

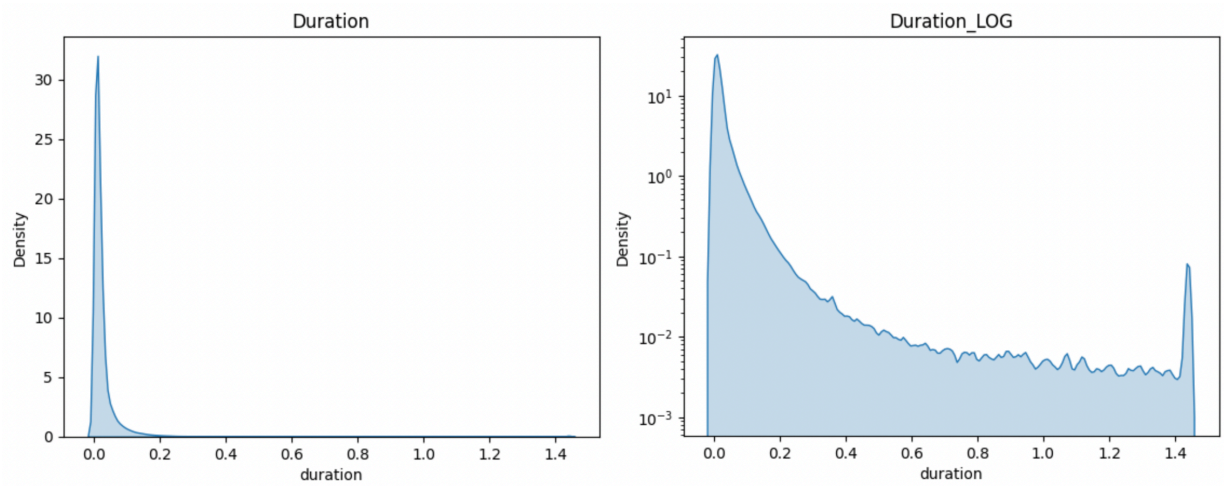
Prueba técnica data science

Estadísticas generales

Classes without Null values:



Distribution of trips by duration variable (linear & log):



Counts of plan duration by minutes and counts of type of passes by trip:

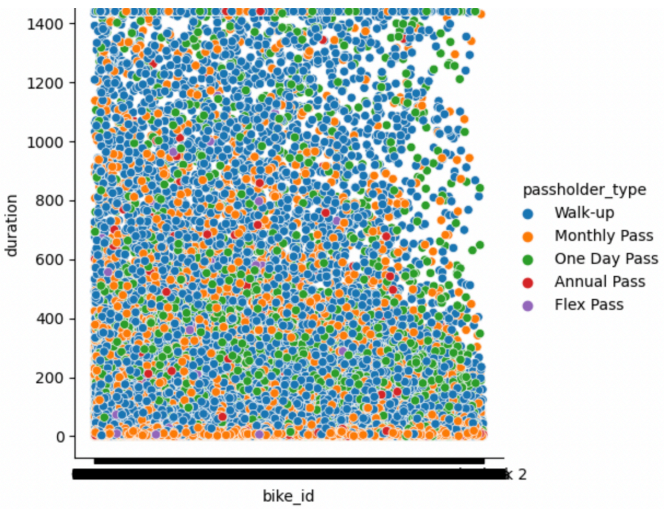
	plan_duration	count
0	1.0	174132
1	0.0	61816
2	30.0	389151
3	365.0	50328
4	150.0	156
5	999.0	43

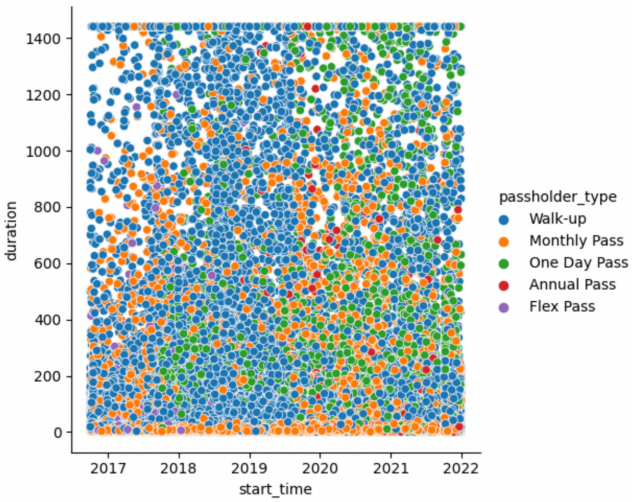
passholder_type	
Annual Pass	32973
Flex Pass	11420
Monthly Pass	387435
One Day Pass	39888
Testing	43
Walk-up	203867

General statistics for types of pass:

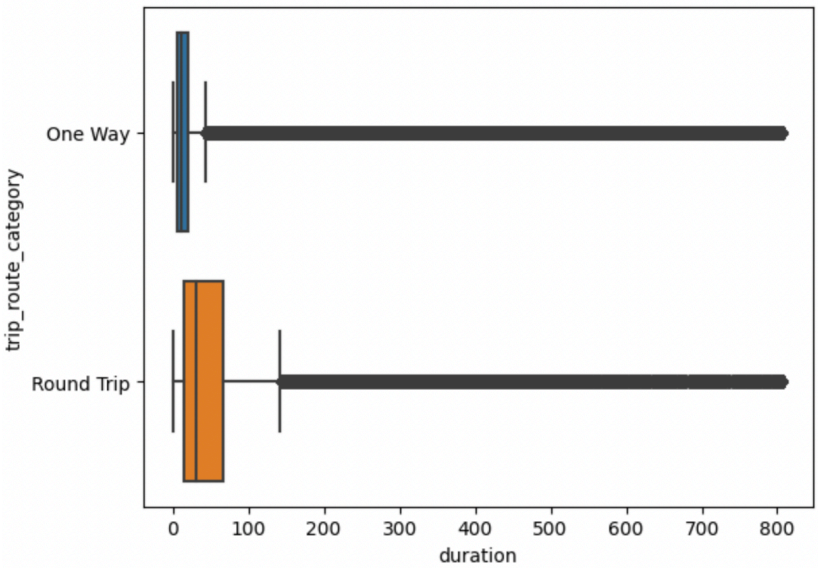
	count	mean	std	min	25%	50%	75%	max
passholder_type								
Annual Pass	32973.0	14.389015	50.602887	1.0	5.0	8.0	14.0	1440.0
Flex Pass	11420.0	16.029159	55.092246	1.0	6.0	9.0	14.0	1440.0
Monthly Pass	387279.0	16.271391	49.415735	1.0	6.0	9.0	17.0	1440.0
One Day Pass	39888.0	73.172257	147.859371	1.0	15.0	29.0	75.0	1440.0
Walk-up	203867.0	54.571804	132.770235	1.0	13.0	25.0	51.0	1440.0

Duration distributions by trips and day/year shows that there is not a clear correlation between those variables, presenting a broad noise pattern.

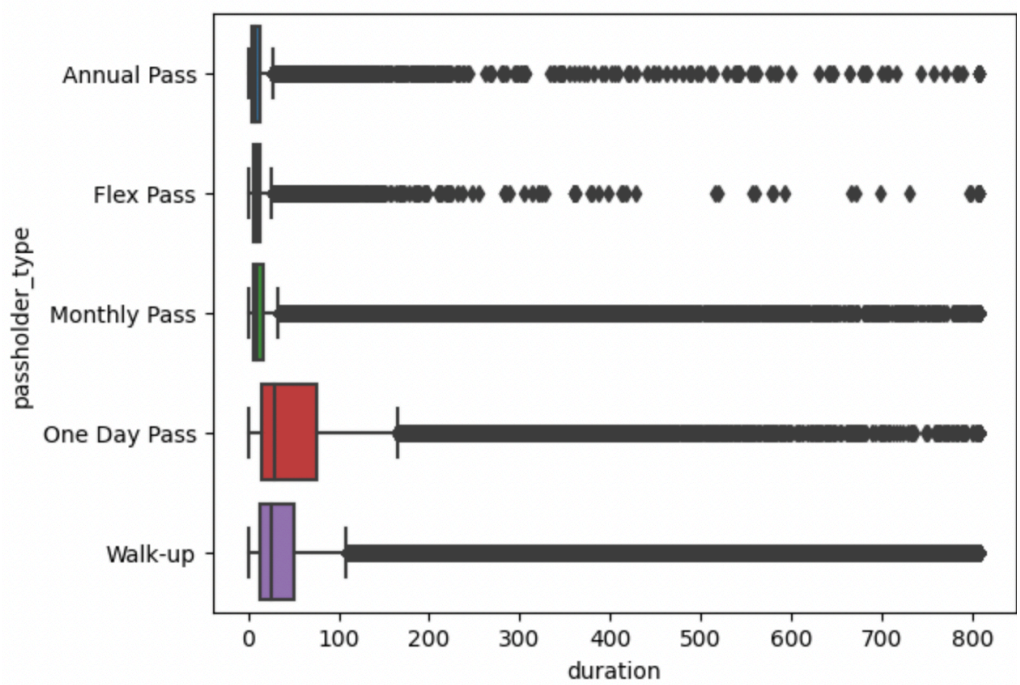




Outliers by trip route categories

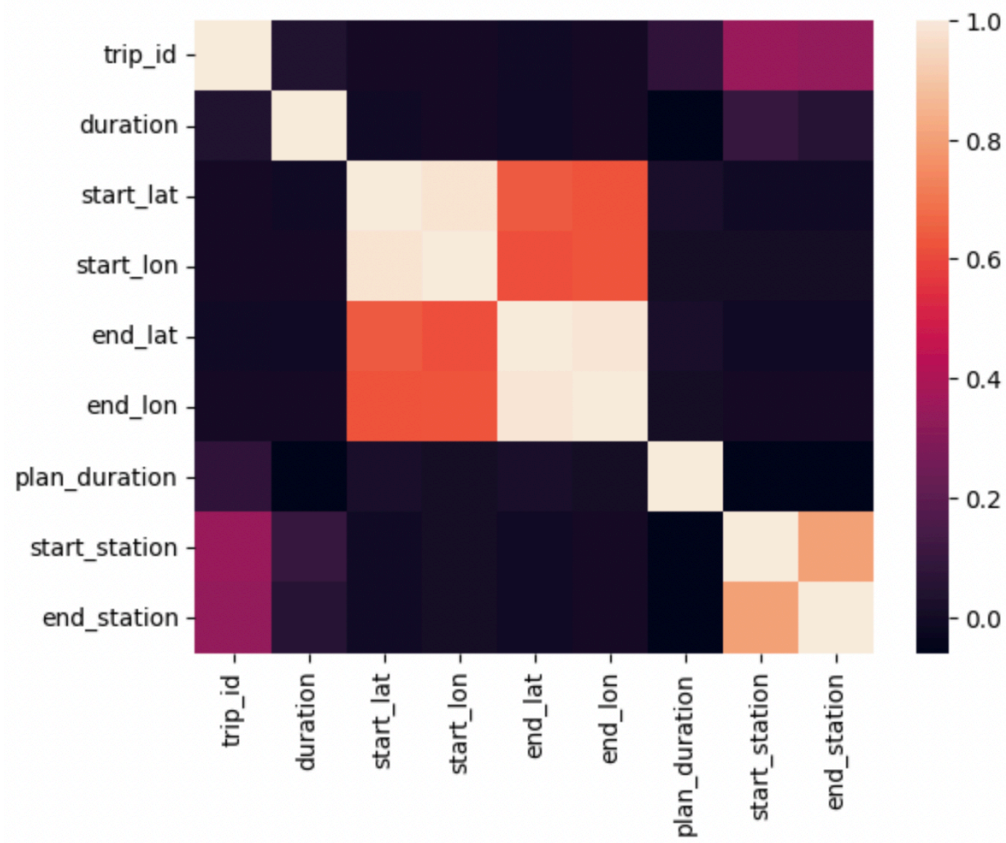


Outliers by pass types regarding the duration of the trips



Correlation of variables:

	trip_id	duration	start_lat	start_lon	end_lat	end_lon	plan_duration	start_station	end_station
trip_id	1.000000	0.034878	-0.000683	0.001853	-0.002211	0.000447	0.073506	0.348450	0.339619
duration	0.034878	1.000000	-0.005112	0.000859	-0.008140	-0.000015	-0.060064	0.097333	0.054521
start_lat	-0.000683	-0.005112	1.000000	0.983237	0.635684	0.622860	0.014488	-0.003942	-0.004703
start_lon	0.001853	0.000859	0.983237	1.000000	0.616067	0.623930	0.004952	0.002344	0.002347
end_lat	-0.002211	-0.008140	0.635684	0.616067	1.000000	0.988989	0.015360	-0.007832	-0.006811
end_lon	0.000447	-0.000015	0.622860	0.623930	0.988989	1.000000	0.005920	-0.001386	-0.001230
plan_duration	0.073506	-0.060064	0.014488	0.004952	0.015360	0.005920	1.000000	-0.055573	-0.057813
start_station	0.348450	0.097333	-0.003942	0.002344	-0.007832	-0.001386	-0.055573	1.000000	0.802540
end_station	0.339619	0.054521	-0.004703	0.002347	-0.006811	-0.001230	-0.057813	0.802540	1.000000



Station Queries

Find the most demanded stations and rank them (top 60):

start_station	station_total_trips				
3005	27909	3008	8897	3016	5085
3030	20066	3075	8822	4212	4810
3014	18469	3034	8749	4220	4472
4214	17111	3038	8723	3040	4430
3031	16790	3027	8424	3046	4327
3035	15065	3032	8308	4211	4323
3064	14470	3047	7934	3055	4305
3082	13451	3052	7836	3033	4065
4210	13125	3019	7210	4345	3934
3006	12413	3048	7103	3028	3776
3069	12385	3076	6541	4491	3670
3042	12357	3066	6425	3058	3628
3067	11492	3062	6400	3036	3509
3007	10633	3018	6237	3078	3422
4215	10504	3029	6203	4202	3306
3068	10261	3026	5650	3056	3300
3074	10237	3051	5555	3024	3292
3022	9793	3023	5336	3037	3239
3049	9562	3011	5285	4346	3106
3063	9156	4209	5112	3077	2956

Find the most demanded stations by month and rank them (top 60) :

start_time_month	start_station	trips_by_month
10	3005	3298
8	3005	2916
11	3005	2681
7	3005	2675
10	3030	2314
6	3005	2302
9	3005	2274
10	3014	2183
1	3005	2107
3	3005	2092
12	3005	2036
2	3005	1997
10	3031	1991
8	4214	1982
10	4214	1908
5	3005	1876
8	3030	1874
11	3030	1860
1	3030	1722
8	3014	1703
12	3030	1656
4	3005	1655
6	3030	1628
5	3035	1625
6	3014	1605
9	3030	1602
11	3014	1594
10	3064	1592
7	4214	1589
10	3035	1586
2	3030	1573
11	3031	1571
5	3014	1569
7	3014	1567
5	3030	1564
7	3030	1559
8	3031	1547
11	4214	1535
10	3069	1513
9	4214	1498
12	3014	1493
3	3030	1488
10	3082	1457
6	3035	1451
9	3014	1441
7	3035	1438
10	3042	1431
	3006	1428
2	3031	1397
12	4214	1392
1	3014	1391
10	4210	1390
2	3014	1388
3	3014	1383
11	3035	1375
3	3031	1367
12	3031	1362
11	3064	1348
7	3031	1330
8	4210	1321

Top 60 trips by station and day of the week:

start_station	start_time_day_of_week	trips_by_day
3005	3	4463
	4	4389
	2	4297
	1	4270
	0	4075
3030	3	3987
	1	3927
	2	3866
3014	3	3560
	1	3488
3005	5	3352
4214	6	3319
3014	2	3296
3030	4	3281
4214	5	3278
3030	0	3197
3005	6	3063
3014	0	2946
	4	2887
3031	4	2668
4210	5	2555
4214	4	2529
4210	6	2527
3031	2	2489
	1	2488
3035	4	2424
3031	3	2416
3082	5	2345
3031	5	2322
	0	2287
3035	3	2240
	2	2218
4215	5	2201
3064	1	2184
4214	0	2167
3035	1	2165
3064	4	2163
3069	5	2154
3064	3	2146
3031	6	2120
3035	0	2117
3064	0	2109
3082	6	2088
3064	2	2087
3069	6	2079
4214	1	2042
3069	4	2035
3042	1	2028
4214	3	2001
3042	4	1986
3035	5	1986
3082	4	1974
3042	3	1954
3064	6	1921
3035	6	1915
4210	4	1912
4215	6	1911
3007	1	1898
3042	2	1863
3006	5	1863

Counts for the most demanded day/hours of the week by station (Top 60):

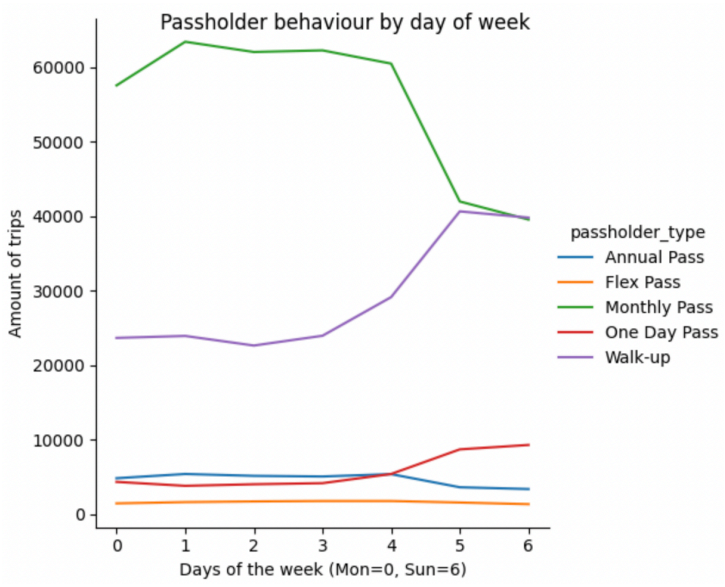
start_station	start_time_day_of_week	start_time_hour	trips_by_hour
3030	3	16	857
	1	16	841
	2	16	772
3014	1	7	750
3030	0	16	739
3014	3	7	715
	2	7	656
	1	8	639
	0	7	625
	2	8	617
3030	1	17	611
	2	17	610
3014	4	7	603
	3	8	587
3030	3	17	582
	1	15	563
3014	1	6	558
3030	3	15	550
	2	15	544
3005	3	18	521
3014	2	6	520
	0	8	511
	3	6	504
3005	1	18	496
3030	4	15	495
3005	2	18	494
3030	4	16	484
3005	0	18	483
3014	4	8	479
3005	1	19	472
3014	0	6	451
3030	0	17	445
3005	0	19	432
	2	17	431
3042	1	7	427
3030	0	15	422
3005	4	18	422
	1	17	412
	3	17	406
		19	405
	2	19	405
4214	6	16	398
3005	4	17	394
4214	6	14	387
3030	4	17	371
4214	6	15	364
	5	17	361
		14	361
3042	2	7	357
4214	5	16	352
		13	352
3005	4	19	349
4214	6	17	345
	5	15	340
3005	0	17	339
3014	4	6	338
3005	3	20	325
3042	3	7	323
3005	4	12	321
3030	2	14	321

Counts for the most demanded day/hours by week and pass type:

start_station	start_time_day_of_week	start_time_hour	passholder_type	trips_by_hour
3030	1	16	Monthly Pass	802
	3	16	Monthly Pass	801
	2	16	Monthly Pass	730
	0	16	Monthly Pass	705
3014	1	7	Monthly Pass	704
	3	7	Monthly Pass	663
	2	7	Monthly Pass	613
	0	7	Monthly Pass	599
	4	7	Monthly Pass	563
	1	8	Monthly Pass	557
3030	1	17	Monthly Pass	553
3014	2	8	Monthly Pass	545
3030	2	17	Monthly Pass	535
	1	15	Monthly Pass	532
	3	17	Monthly Pass	521
3014	3	8	Monthly Pass	512
	1	6	Monthly Pass	498
3030	3	15	Monthly Pass	497
	2	15	Monthly Pass	479
3014	0	8	Monthly Pass	461
3030	4	15	Monthly Pass	456
3014	2	6	Monthly Pass	456
	3	6	Monthly Pass	454
3030	4	16	Monthly Pass	443
3014	4	8	Monthly Pass	420
3030	0	17	Monthly Pass	407
3014	0	6	Monthly Pass	406
3005	2	18	Monthly Pass	404
3042	1	7	Monthly Pass	403
3005	3	18	Monthly Pass	398
	1	18	Monthly Pass	389
	0	18	Monthly Pass	388
3030	0	15	Monthly Pass	388
3005	1	19	Monthly Pass	352
	2	17	Monthly Pass	344
	4	18	Monthly Pass	340
3042	2	7	Monthly Pass	335
3005	1	17	Monthly Pass	324
	3	19	Monthly Pass	324
4214	6	16	Walk-up	323
3005	0	19	Monthly Pass	321
	2	19	Monthly Pass	319
4214	6	14	Walk-up	310
3005	3	17	Monthly Pass	308
3030	4	17	Monthly Pass	308
3005	4	17	Monthly Pass	308
3042	3	7	Monthly Pass	302
4214	5	17	Walk-up	291
		14	Walk-up	290
	6	15	Walk-up	289
		17	Walk-up	287
3014	4	6	Monthly Pass	286
3030	2	14	Monthly Pass	285
4214	5	13	Walk-up	284
		16	Walk-up	281
3005	0	17	Monthly Pass	277
3042	1	8	Monthly Pass	276
	0	7	Monthly Pass	274
4214	5	15	Walk-up	266
3030	1	14	Monthly Pass	263

Pass type behaviour by day of the week:

passholder_type	start_time_day_of_week	trip_count
Annual Pass	0	4853
	1	5422
	2	5171
	3	5084
	4	5402
	5	3640
	6	3401
Flex Pass	0	1480
	1	1649
	2	1738
	3	1794
	4	1793
	5	1595
	6	1371
Monthly Pass	0	57556
	1	63410
	2	62037
	3	62255
	4	60480
	5	41981
	6	39560
One Day Pass	0	4362
	1	3832
	2	4046
	3	4189
	4	5420
	5	8724
	6	9315
Walk-up	0	23679
	1	23942
	2	22654
	3	23967
	4	29153
	5	40654
	6	39818



Monthly passes are used mainly for weekdays trips, while walk-up passes are used mainly on the weekends.

We can see a similar behaviour between one day passes and Annual pass but with much less variation.

The use of Flex passes are almost uniform between the weekends and the weekdays.

Most frequent travels between stations:

	start_station	end_station	Freq
43675	4214	4214	7930
7328	3030	3014	6123
42207	4210	4210	6067
2583	3014	3030	5364
43676	4214	4215	4528
44042	4215	4215	3620
0	3005	3005	3152
42212	4210	4215	2944
63126	4345	4345	2522
7686	3031	3005	2376
43671	4214	4210	2222
42211	4210	4214	2217
22	3005	3031	2126
24224	4126	4126	1901
113403	4543	4543	1894
44041	4215	4214	1795
9176	3035	3035	1720
7352	3030	3042	1715
63493	4346	4346	1713
2571	3014	3016	1568
7708	3031	3031	1488
42574	4211	4211	1470
11367	3042	3030	1450
42941	4212	4212	1442
45143	4220	4220	1416
113770	4544	4544	1412
367	3006	3006	1370
23121	3082	3082	1370
2936	3016	3014	1365
17616	3063	3063	1313
44037	4215	4210	1293
62759	4344	4344	1279
31564	4153	4153	1256
41840	4209	4209	1198
2570	3014	3014	1188
12846	3048	3048	1172
7341	3030	3030	1161
42944	4212	4215	1150
19818	3069	3069	1140
20185	3074	3074	1114
41846	4209	4215	1104
13908	3052	3005	1099
19032	3067	3005	1078
18717	3066	3066	1071
56	3005	3075	1070
12479	3047	3047	1068
49	3005	3064	1057
8810	3034	3035	1041
11378	3042	3042	1041
19084	3067	3067	1037
9175	3035	3034	1022
17983	3064	3064	1018
17956	3064	3031	1004

Travels between stations gives another perspective of demand of trips by station and presents a general flow of the bikers and how the travel through the city.

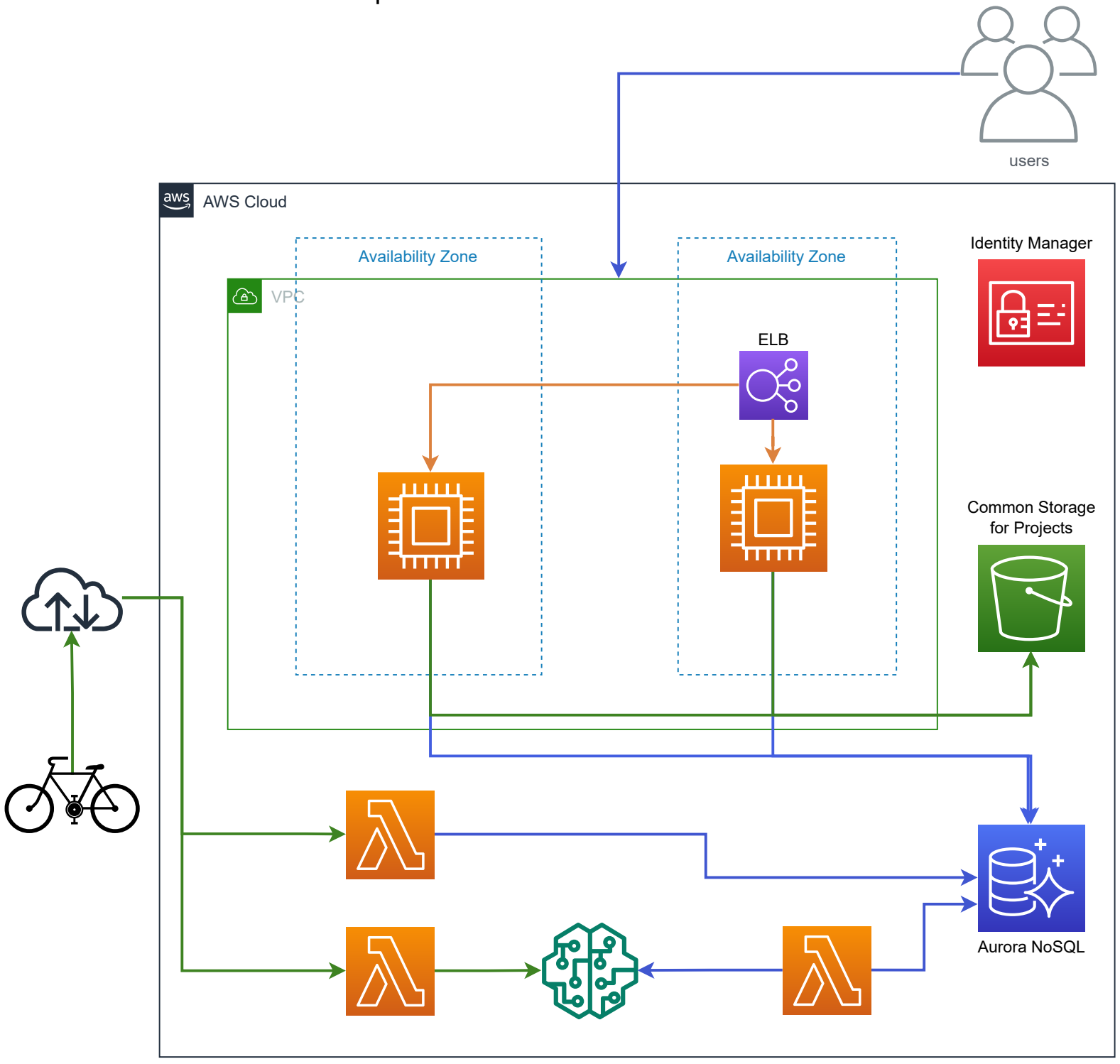
With more insights between these trajectories it could be planned where more stations should be placed to favour a more efficient city paths and station planning.

Amount of trips by quarter and most used pass types (annually) by station:

	trips_by_quarter				total_trips	most_used_pass_type_by_year
start_time_quarter	1	2	3	4		
start_station						
3005	6196	5833	7865	8015	27909	Monthly Pass
3030	4783	4418	5035	5830	20066	Monthly Pass
3014	4162	4326	4711	5270	18469	Monthly Pass
4214	3496	3711	5069	4835	17111	Walk-up
3031	4080	3615	4171	4924	16790	Monthly Pass
3035	3298	4364	3333	4070	15065	Monthly Pass
3064	3286	3565	3584	4035	14470	Monthly Pass
3082	3019	3449	3342	3641	13451	Monthly Pass
4210	2742	2966	3728	3689	13125	Walk-up
3006	2581	2757	3612	3463	12413	Monthly Pass
3069	2714	2945	3109	3617	12385	Monthly Pass
3042	2806	3110	2985	3456	12357	Monthly Pass
3067	2457	2962	2827	3246	11492	Monthly Pass
3007	2254	2516	2836	3027	10633	Monthly Pass
4215	2038	2402	2977	3087	10504	Walk-up
3068	2248	2607	2686	2720	10261	Monthly Pass
3074	2156	2536	2687	2858	10237	Monthly Pass
3022	2330	2166	2336	2961	9793	Monthly Pass
3049	2249	2113	2277	2923	9562	Monthly Pass
3063	2146	2253	1893	2864	9156	Monthly Pass
3008	1891	2120	2367	2519	8897	Monthly Pass
3075	1875	2096	2348	2503	8822	Monthly Pass
3034	1997	2022	2209	2521	8749	Monthly Pass
3038	1742	2015	2358	2608	8723	Monthly Pass
3027	1943	1979	2150	2352	8424	Monthly Pass
3032	1839	1812	2287	2370	8308	Monthly Pass
3047	1768	1933	2023	2210	7934	Monthly Pass
3052	1724	1940	1880	2292	7836	Monthly Pass
3019	1461	1741	1941	2067	7210	Monthly Pass
3048	1682	1721	1640	2060	7103	Monthly Pass
3076	1580	1711	1584	1666	6541	Monthly Pass
3066	1462	1615	1609	1739	6425	Monthly Pass
3062	1481	1437	1538	1944	6400	Monthly Pass
3018	1451	1504	1622	1660	6237	Monthly Pass
3029	1469	1474	1512	1748	6203	Monthly Pass
3026	1218	1335	1457	1640	5650	Monthly Pass
3051	1321	1167	1296	1771	5555	Monthly Pass
3023	1202	1337	1386	1411	5336	Monthly Pass
3011	1271	1242	1385	1387	5285	Monthly Pass
4209	1153	1304	1434	1221	5112	Walk-up
3016	1144	1227	1291	1423	5085	Monthly Pass
4212	1090	1162	1249	1309	4810	Walk-up
4220	874	1049	1278	1271	4472	Monthly Pass
3040	921	1120	1180	1209	4430	Monthly Pass
3046	973	934	1111	1309	4327	Monthly Pass
4211	1002	922	1141	1258	4323	Walk-up
3055	849	1045	1255	1156	4305	Monthly Pass
3033	1079	922	1001	1063	4065	Monthly Pass
4345	620	1269	1301	744	3934	Walk-up
3028	783	1030	950	1013	3776	Monthly Pass
4491	960	1137	661	912	3670	Monthly Pass
3058	863	813	912	1040	3628	Monthly Pass
3036	775	893	864	977	3509	Monthly Pass
3078	828	781	837	976	3422	Monthly Pass
4202	666	764	918	958	3306	Walk-up
3056	772	828	736	964	3300	Monthly Pass
3024	687	765	864	976	3292	Monthly Pass
3037	708	787	841	903	3239	Monthly Pass
4346	579	871	1095	561	3106	Walk-up
3077	675	721	759	801	2956	Monthly Pass

Top 5 demanded stations and majority of plans that demand trips from there:

		trip_id
start_station	passholder_type	
3005	Monthly Pass	20423
3030	Monthly Pass	15953
4214	Walk-up	13361
3014	Monthly Pass	13198
3031	Monthly Pass	12333

Puesta en producción de modelo

- Serverless categorization + data exploration
- Serverless model and db update