

Data Analytics in the Aviation Industry

Understand your flight data

```
display.Image('/Users/Yedou/Documents/Flight-analytics-from-ge-digital-3200x1404.jpg', width = 1200, height = 300)
```



Data Engineering

Using data engineering to combine full-flight engine data, part manufacturing data, airport location data to create a data lake our analysts can use

Postgres/SQL | Data orchestration & transformation | Critical Thinking

```
import pandas as pd
import numpy as np
#import gspread as gs
import matplotlib.pyplot as plt
from IPython import display
import matplotlib.colors as mcolors
import folium
import json

df_engine_axm = pd.read_excel('/Users/Yedou/Documents/GE_Dataset_Task1 (1) (2).xlsx',
sheet_name='av_engine_data_axm_psql')

df_engine_axm
```

	dataset	esn	unit	flight_cycle	datetime
operator \					
0	test_FD001	999126.0	26.0	1.0	2017-12-31 18:33:07
AXM					
1	test_FD001	999126.0	26.0	2.0	2017-12-31 19:59:49

AXM							
2	test_FD001	999126.0	26.0	3.0	2017-12-31	21:44:38	
AXM							
3	test_FD001	999126.0	26.0	4.0	2018-01-01	00:50:02	
AXM							
4	test_FD001	999126.0	26.0	5.0	2018-01-01	04:25:32	
AXM							
...
...							
21183	train_FD001	999096.0	96.0	264.0	2018-02-13	22:32:30	
AXM							
21184	train_FD001	999096.0	96.0	265.0	2018-02-14	00:48:00	
AXM							
21185	train_FD001	999096.0	96.0	266.0	2018-02-14	02:06:00	
AXM							
21186	train_FD001	999096.0	96.0	267.0	2018-02-14	03:23:32	
AXM							
21187	train_FD001	999096.0	96.0	268.0	2018-02-14	17:04:38	
AXM							

	depart_icao	destination_icao		hpc_eff_mod	hpc_flow_mod	...
phi \						
0	VTBD	VTUV		-0.0027	0.0006	...
522.55						
1	VTUV	VTBD		-0.0029	0.0002	...
521.85						
2	VTBD	VMMC		0.0008	0.0001	...
522.34						
3	VMMC	VTBD		-0.0026	0.0005	...
522.03						
4	VTBD	VTSP		0.0020	0.0005	...
522.04						
...
...						
21183	WMKK	VDSV		-0.0034	0.0005	...
519.47						
21184	VDSV	WMKK		0.0015	0.0004	...
519.49						
21185	WMKK	WBKL		-0.0028	-0.0002	...
519.48						
21186	WMKK	WBKL		0.0001	0.0005	...
519.61						
21187	WMKK	VDSR		-0.0031	-0.0005	...
519.18						

	nrf	nrc	bpr	farb	htbleed	nf_dmd	pcnfr_dmd
w31 \							
0	2387.98	8147.17	8.3468	0.03	392.0	2388.0	100.0
38.82							
1	2388.01	8146.58	8.4159	0.03	390.0	2388.0	100.0

39.00							
2	2388.05	8144.60	8.4444	0.03	391.0	2388.0	100.0
39.20							
3	2388.00	8147.98	8.3986	0.03	392.0	2388.0	100.0
38.97							
4	2388.00	8142.07	8.4332	0.03	391.0	2388.0	100.0
39.11							
...
21183	2388.25	8142.08	8.5109	0.03	395.0	2388.0	100.0
38.53							
21184	2388.20	8140.98	8.5158	0.03	396.0	2388.0	100.0
38.30							
21185	2388.27	8145.65	8.5165	0.03	395.0	2388.0	100.0
38.63							
21186	2388.25	8143.74	8.5615	0.03	396.0	2388.0	100.0
38.51							
21187	2388.24	8144.03	8.5022	0.03	397.0	2388.0	100.0
38.36							

	w32
0	23.3876
1	23.3793
2	23.3390
3	23.3931
4	23.3904
...	...
21183	23.0965
21184	23.0309
21185	23.1059
21186	23.0118
21187	23.0258

[21188 rows x 32 columns]

```
df_engine_fron = pd.read_excel('/Users/Yedou/Documents/GE_Dataset_Task1 (1) (2).xlsx',
sheet_name='av_engine_data_fron_psql')
```

df_engine_fron

	dataset	esn	unit	flight_cycle	datetime
operator \					
0	train_FD001	999050.0	50.0	1.0	2018-01-06 06:01:09
FRON					
1	train_FD001	999050.0	50.0	2.0	2018-01-06 07:41:00
FRON					
2	train_FD001	999050.0	50.0	3.0	2018-01-06 08:41:18
FRON					
3	train_FD001	999050.0	50.0	4.0	2018-01-06 10:14:00
FRON					

4	train_FD001	999050.0	50.0	5.0	2018-01-06	11:12:52
FRON						
...
...						
7285	train_FD001	999086.0	86.0	271.0	2018-02-08	08:24:59
FRON						
7286	train_FD001	999086.0	86.0	272.0	2018-02-08	13:16:00
FRON						
7287	train_FD001	999086.0	86.0	273.0	2018-02-08	14:16:02
FRON						
7288	train_FD001	999086.0	86.0	274.0	2018-02-08	18:21:45
FRON						
7289	train_FD001	999086.0	86.0	275.0	2018-02-09	09:05:00
FRON						

	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	...
phi \					
0	KMCO	KMSY	-0.0029	-0.0002	...
521.68					
1	KMSY	KSAT	-0.0002	-0.0005	...
522.12					
2	KMSY	KSAT	-0.0010	-0.0005	...
521.96					
3	KSAT	KSAN	-0.0061	-0.0002	...
522.31					
4	KSAT	KSAN	-0.0002	0.0001	...
522.05					
...
...					
7285	KMIA	KLAS	-0.0017	0.0000	...
519.80					
7286	KLAS	KIND	0.0002	-0.0004	...
519.62					
7287	KLAS	KIND	0.0017	0.0002	...
519.51					
7288	KIND	KMCO	0.0003	0.0003	...
519.13					
7289	KMCO	KBUF	-0.0002	-0.0005	...
519.63					

	nrf	nrc	bpr	farb	htbleed	nf_dmd	pcnfr_dmd
w31 \							
0	2388.06	8151.49	8.4158	0.03	393.0	2388.0	100.0
38.80							
1	2388.07	8142.72	8.4467	0.03	392.0	2388.0	100.0
38.99							
2	2388.05	8139.14	8.4424	0.03	393.0	2388.0	100.0
38.91							
3	2388.04	8145.16	8.4504	0.03	393.0	2388.0	100.0
38.95							

4	2388.05	8145.35	8.3822	0.03	392.0	2388.0	100.0
38.83							
...
.							
7285	2388.22	8153.15	8.5043	0.03	396.0	2388.0	100.0
38.47							
7286	2388.23	8153.33	8.4910	0.03	396.0	2388.0	100.0
38.52							
7287	2388.21	8150.99	8.5339	0.03	396.0	2388.0	100.0
38.45							
7288	2388.25	8157.08	8.5220	0.03	396.0	2388.0	100.0
38.49							
7289	2388.22	8151.89	8.4919	0.03	398.0	2388.0	100.0
38.51							

	w32
0	23.3016
1	23.3440
2	23.3190
3	23.3161
4	23.3256
...	...
7285	23.0271
7286	23.0738
7287	23.1261
7288	23.0981
7289	22.9947

[7290 rows x 32 columns]

```
df_engine_aic = pd.read_excel('/Users/Yedou/Documents/GE_Dataset_Task1 (1) (2).xlsx',
sheet_name='av_engine_data_aic_psql')
```

df_engine_aic

	dataset	esn	unit	flight_cycle	datetime
operator \					
0	train_FD001	999001.0	1.0	1.0	2018-01-11 17:10:56
AIC					
1	train_FD001	999001.0	1.0	2.0	2018-01-12 01:07:00
AIC					
2	train_FD001	999001.0	1.0	3.0	2018-01-12 02:06:59
AIC					
3	train_FD001	999001.0	1.0	4.0	2018-01-12 07:42:00
AIC					
4	train_FD001	999001.0	1.0	5.0	2018-01-12 08:42:17
AIC					
...
...					
5815	train_FD001	999069.0	69.0	330.0	2018-02-17 19:08:17

AIC							
5816	train_FD001	999069.0	69.0	331.0	2018-02-17	22:41:34	
AIC							
5817	train_FD001	999069.0	69.0	332.0	2018-02-18	04:41:00	
AIC							
5818	train_FD001	999069.0	69.0	333.0	2018-02-18	05:41:31	
AIC							
5819	train_FD001	999069.0	69.0	334.0	2018-02-18	08:05:00	
AIC							

	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	...
phi \					
0	VTBS	VICG	-0.0005	0.0002	...
521.93					
1	VICG	VTBS	0.0016	-0.0003	...
521.72					
2	VICG	VTBS	-0.0017	0.0001	...
521.97					
3	VTBS	VIDP	-0.0012	-0.0002	...
522.04					
4	VTBS	VIDP	0.0027	0.0001	...
522.25					
...
...					
5815	VIDP	VOMM	0.0040	-0.0002	...
520.31					
5816	VOMM	VIDP	0.0013	0.0003	...
519.86					
5817	VIDP	VECC	0.0004	-0.0001	...
518.69					
5818	VIDP	VECC	0.0016	-0.0002	...
520.48					
5819	VECC	VIDP	0.0007	0.0000	...
519.12					

	nrf	nrc	bpr	farb	htbleed	nf_dmd	pcnfr_dmd	
w31 \								
0	2388.12	8130.36	8.4279	0.03	392.0	2388.0	100.0	
38.86								
1	2388.09	8134.28	8.4230	0.03	391.0	2388.0	100.0	
38.80								
2	2388.09	8133.95	8.3951	0.03	391.0	2388.0	100.0	
38.84								
3	2388.10	8129.85	8.4040	0.03	391.0	2388.0	100.0	
39.03								
4	2387.97	8132.67	8.4135	0.03	392.0	2388.0	100.0	
38.80								
...
.								
5815	2388.26	8129.59	8.5460	0.03	397.0	2388.0	100.0	

```

38.53
5816 2388.27 8131.20 8.5116 0.03 396.0 2388.0 100.0
38.28
5817 2388.26 8126.63 8.5059 0.03 396.0 2388.0 100.0
38.47
5818 2388.32 8132.97 8.5270 0.03 396.0 2388.0 100.0
38.52
5819 2388.33 8124.27 8.5126 0.03 397.0 2388.0 100.0
38.42

```

```

          w32
0      23.4513
1      23.3881
2      23.3085
3      23.3975
4      23.3820
...
5815 23.0549
5816 23.0610
5817 23.0422
5818 23.1208
5819 22.9782

```

[5820 rows x 32 columns]

```

df_engine_pgt = pd.read_excel('/Users/Yedou/Documents/GE_Dataset_
Task1 (1) (2).xlsx',
sheet_name='av_engine_data_pgt_psql')

```

```
df_engine_pgt
```

```

          dataset      esn  unit  flight_cycle      datetime
operator \
0      train_FD001  999056.0  56.0           1.0 2018-01-01 06:33:13
PGT
1      train_FD001  999056.0  56.0           2.0 2018-01-01 09:40:21
PGT
2      train_FD001  999056.0  56.0           3.0 2018-01-01 12:23:01
PGT
3      train_FD001  999056.0  56.0           4.0 2018-01-01 14:11:10
PGT
4      train_FD001  999056.0  56.0           5.0 2018-01-01 21:10:50
PGT
...          ...      ...      ...          ...          ...
...
6295  train_FD001  999084.0  84.0          257.0 2018-06-06 08:22:54
PGT
6296  train_FD001  999084.0  84.0          258.0 2018-06-06 12:01:39
PGT
6297  train_FD001  999084.0  84.0          259.0 2018-06-06 14:01:52
PGT

```

6298	train_FD001	999084.0	84.0	260.0	2018-06-06	18:41:30
PGT						
6299	train_FD001	999095.0	95.0	266.0	2018-02-08	01:30:23
PGT						

	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	...
phi \					
0	LTBJ	LTCR	0.0012	-0.0004	...
521.18					
1	LTCR	LTBJ	0.0012	-0.0004	...
520.82					
2	LTBJ	LTFJ	0.0026	0.0005	...
521.70					
3	LTFJ	LTCG	0.0034	-0.0002	...
522.06					
4	LTCG	LTFJ	0.0024	-0.0001	...
521.36					
...
...					
6295	LFSB	LTFJ	-0.0027	0.0000	...
519.66					
6296	LTFJ	LTAJ	0.0026	0.0001	...
520.00					
6297	LTAJ	LTFJ	-0.0013	0.0000	...
519.74					
6298	OKBK	LTFJ	-0.0023	0.0001	...
519.77					
6299	LTFJ	EDDS	0.0019	0.0003	...
519.27					

	nrf	nrc	bpr	farb	htbleed	nf_dmd	pcnfr_dmd	
w31 \								
0	2388.13	8136.92	8.4412	0.03	395.0	2388.0	100.0	
38.81								
1	2388.08	8133.11	8.4461	0.03	394.0	2388.0	100.0	
38.82								
2	2388.12	8136.86	8.4357	0.03	394.0	2388.0	100.0	
38.89								
3	2388.18	8132.53	8.4411	0.03	394.0	2388.0	100.0	
38.79								
4	2388.16	8133.47	8.4824	0.03	394.0	2388.0	100.0	
39.00								
...
.								
6295	2388.20	8148.87	8.5464	0.03	397.0	2388.0	100.0	
38.49								
6296	2388.18	8148.76	8.5113	0.03	396.0	2388.0	100.0	
38.49								
6297	2388.20	8146.22	8.5336	0.03	395.0	2388.0	100.0	
38.36								

6298	2388.23	8149.66	8.5367	0.03	396.0	2388.0	100.0
38.49							
6299	2388.20	8183.58	8.5648	0.03	396.0	2388.0	100.0
38.37							

	w32
0	23.2391
1	23.3340
2	23.2844
3	23.3204
4	23.3592
...	...
6295	22.9996
6296	23.0383
6297	23.0022
6298	23.1018
6299	23.1041

[6300 rows x 32 columns]

provide list of dataframes

```
df_engine_concat =
pd.concat([df_engine_axm,df_engine_fron,df_engine_aic,df_engine_pgt])
```

df_engine_concat

	dataset	esn	unit	flight_cycle	datetime
operator \					
0	test_FD001	999126.0	26.0	1.0	2017-12-31 18:33:07
AXM					
1	test_FD001	999126.0	26.0	2.0	2017-12-31 19:59:49
AXM					
2	test_FD001	999126.0	26.0	3.0	2017-12-31 21:44:38
AXM					
3	test_FD001	999126.0	26.0	4.0	2018-01-01 00:50:02
AXM					
4	test_FD001	999126.0	26.0	5.0	2018-01-01 04:25:32
AXM					
...
...					
6295	train_FD001	999084.0	84.0	257.0	2018-06-06 08:22:54
PGT					
6296	train_FD001	999084.0	84.0	258.0	2018-06-06 12:01:39
PGT					
6297	train_FD001	999084.0	84.0	259.0	2018-06-06 14:01:52
PGT					
6298	train_FD001	999084.0	84.0	260.0	2018-06-06 18:41:30
PGT					
6299	train_FD001	999095.0	95.0	266.0	2018-02-08 01:30:23
PGT					

	w32
0	23.3876
1	23.3793
2	23.3390
3	23.3931
4	23.3904
...	...
6295	22.9996
6296	23.0383
6297	23.0022
6298	23.1018
6299	23.1041

[40598 rows x 32 columns]

```
df_manuf_psql = pd.read_excel('/Users/Yedou/Documents/GE_Dataset_Task1 (1) (2).xlsx',
sheet_name='av_bom_manufacturing_psql')
df_manuf_psql
```

	esn	pn	sn	desc	vstream
0	999010.0	54321P01	8.221064e+08	shroud	cmc
1	999010.0	54321P01	6.644757e+08	shroud	cmc
2	999010.0	54321P01	2.430976e+09	shroud	cmc
3	999010.0	54321P01	1.277358e+09	shroud	cmc
4	999010.0	54321P01	8.668055e+09	shroud	cmc
...
20195	999093.0	44321P02	1.003440e+09	blade	machined_airfoils
20196	999093.0	44321P02	3.829220e+09	blade	machined_airfoils
20197	999093.0	44321P02	4.571830e+09	blade	machined_airfoils
20198	999093.0	44321P02	8.136479e+09	blade	machined_airfoils
20199	999093.0	65421P11	4.508429e+09	disk	rotating_parts

[20200 rows x 5 columns]

```
df_manuf_supply_chain =
pd.read_excel('/Users/Yedou/Documents/GE_Dataset_Task1 (1) (2).xlsx',
sheet_name='av_manufacturing_supply_chain')
df_manuf_supply_chain
```

	sn	pn	op	part_desc	kc	msmts
max \						
0	7.837606e+09	54321P01	op116	shroud	1.0	31.983503
33.061659						
1	5.039652e+09	54321P01	op116	shroud	1.0	34.456691
33.061659						
2	7.837606e+09	54321P01	op220	shroud	2.0	27.895096
30.303501						


```

...
63995  6.299767e+09  54321P01  op116    shroud  1.0  20.554360
33.061659
63996  4.512062e+09  54321P01  op116    shroud  1.0  22.756896
33.061659
63997  6.299767e+09  54321P01  op220    shroud  2.0  29.583411
30.303501
63998  4.512062e+09  54321P01  op220    shroud  2.0  30.475523
30.303501
63999  4.567413e+09  44321P02  op016    blade   2.0  19.191280
27.987527

```

```

min
0      21.160852
1      21.160852
2      17.044897
3      17.044897
4      10.600079

```

```

...
63995  21.160852
63996  21.160852
63997  17.044897
63998  17.044897
63999  11.183152

```

[64000 rows x 8 columns]

```

inner_join = pd.merge(df_manuf_supply_chain,
                      df_manuf_psql,
                      on = 'sn',
                      how = 'inner')

```

inner_join

```

max \
      sn      pn_x      op part_desc  kc      msmts
0      7.837606e+09  54321P01  op116    shroud  1.0  31.983503
33.061659
1      7.837606e+09  54321P01  op220    shroud  2.0  27.895096
30.303501
2      5.039652e+09  54321P01  op116    shroud  1.0  34.456691
33.061659
3      5.039652e+09  54321P01  op220    shroud  2.0  32.920628
30.303501
4      9.856636e+09  44321P02  op420    blade   1.0  12.640872
16.346054
...      ...      ...      ...      ...      ...
...
63995  5.762950e+09  54321P01  op220    shroud  2.0  26.486150

```

```

30.303501
63996  6.299767e+09  54321P01  op116      shroud  1.0  20.554360
33.061659
63997  6.299767e+09  54321P01  op220      shroud  2.0  29.583411
30.303501
63998  4.512062e+09  54321P01  op116      shroud  1.0  22.756896
33.061659
63999  4.512062e+09  54321P01  op220      shroud  2.0  30.475523
30.303501

```

	min	esn	pn_y	desc	vstream
0	21.160852	999001.0	54321P01	shroud	cmc
1	17.044897	999001.0	54321P01	shroud	cmc
2	21.160852	999001.0	54321P01	shroud	cmc
3	17.044897	999001.0	54321P01	shroud	cmc
4	10.600079	999001.0	44321P02	blade	machined_airfoils
...
63995	17.044897	999199.0	54321P01	shroud	cmc
63996	21.160852	999200.0	54321P01	shroud	cmc
63997	17.044897	999200.0	54321P01	shroud	cmc
63998	21.160852	999200.0	54321P01	shroud	cmc
63999	17.044897	999200.0	54321P01	shroud	cmc

[64000 rows x 12 columns]

```

inner_join.to_excel('/Users/Yedou/Documents/export_inner_join.xlsx',
index = False, header=True)

```

<https://towardsdatascience.com/creating-a-simple-map-with-folium-and-python-4c083abfff94>
<https://campus.datacamp.com/courses/visualizing-geospatial-data-in-python/geoseries-and-folium?ex=11>

Data Visualization

Use data to create a run chart and KPI (key performance indicators) tables based off of simulated Aviation data

Run Chart | Data storytelling through visualizations | Business Intelligence (BI)

```

df_map = pd.read_excel('/Users/Yedou/Documents/GE_Dataset_Task1 (1)
(2).xlsx',
sheet_name='av_lkp_airport_codes_t_psql')

```

```
df_map
```

	airport_icao	destination_latitude	destination_longitude
0	VISR	33.988	74.775

1	VLVT	17.988	102.563
2	VTSG	8.100	98.985
3	KONT	34.056	-117.601
4	MMPR	20.680	-105.254
...
314	LTCE	39.956	41.171
315	VLLB	19.906	102.169
316	VTSS	6.933	100.395
317	WADD	-8.747	115.169
318	WBGB	3.124	113.020

[319 rows x 3 columns]

```
df_map.to_excel('/Users/Yedou/Documents/export_df_map.xlsx', index =
False, header=True)
```

Creating the map

For the map, the first step is to create a map of the location I want. Using the location parameter, I pass in the mean of the latitude and longitude coordinates I have to centre the map there.

```
map = folium.Map(location=[df_map['destination_latitude'].mean(),
df_map['destination_longitude'].mean()], zoom_start=14,
control_scale=True)
```

map

<folium.folium.Map at 0x1248a7df0>

Step 5 — Adding points to the map

Now I add the points for each airport locations to the map. Iterating through each row of the dataframe, I pass the location latitude and longitudes to folium.Marker as a list and pass the name to the popup parameter. And for each location I add to map:

```
for index, location_info in df_map.iterrows():
    folium.Marker([location_info["destination_latitude"],
location_info["destination_longitude"]],
popup=location_info["airport_icao"]).add_to(map)
```

map

<folium.folium.Map at 0x1248a7df0>