

Analysis of SpaceX Launch Data

John Isaac Enriquez

January 20, 2022

Outline

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

Section 1

Executive Summary

Executive Summary

Methodology

The SpaceX launch data were collected and analyzed using various data science methods. Data collection is performed using web scraping and API requests. Preliminary data exploration was performed using SQL queries. Location data were visualized using Folium maps. A dashboard containing selected data features was also developed. The data was then imported and processed in Python Pandas and the features that correlate to successful first stage rocket landing were selected. Four classification methods were used to develop machine learning models that will predict the outcome of every launch.

Results

There are four launch sites, one of which is in Los Angeles and three are in Florida. Location data shows that the launch sites are in close proximity to the coast. The features that correlates to the launch outcome includes the launch date, launch site, payload weight, and orbit type. The developed machine learning models are able to predict the outcome of the launch with accuracy scores of 0.833 for the logistic regression, support vector machine, and K nearest neighbor, and 0.778 for the decision tree.

Section 2

Introduction

Introduction

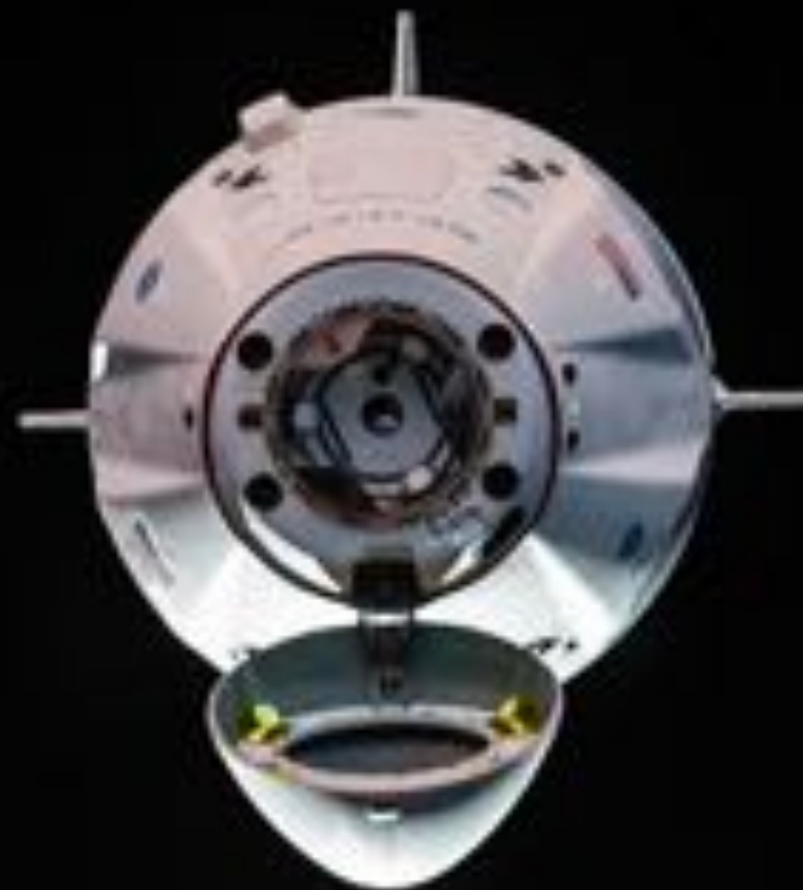
SpaceX is an American aerospace manufacturer, space transportation services, and communications corporation in Hawthorne, California. SpaceX manufactures the Falcon 9 launch vehicle that has a reusable first stage. This feature makes SpaceX Falcon 9 services about million dollar cheaper compared to other providers.

The objective of this project is to predict whether the first stage of Falcon 9 will successfully land. This determines the cost of the launch which is a valuable information for a competitor who wants to bid against SpaceX for a rocket launch. This will be done by gathering, analyzing, and modelling SpaceX launch data such as rocket used, payload delivered, launch specifications, landing specifications, and landing outcome which is available through SpaceX REST API.

The general objective of this project is to answer the following questions:

What features determines the launch outcome?

What machine learning method best predicts the outcome?



Section 3

Methodology

Methodology

Executive Summary

Data collection methodology:

Data was collected using API requests and web scraping

Perform data wrangling

Data was processed using Python Pandas

Perform exploratory data analysis (EDA) using visualization and SQL

EDA was performed using SQL queries and visualized using Matplotlib and Seaborn graphs

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

Classification models were built using Python Scikit Learn module. The data set was divided into training and testing sets. The features were preprocessed using StandardScaler function. The parameters were tuned using GridSearchCV. Finally, the accuracy of each model was evaluated using the score function.

Data Collection

SpaceX launch data were collected from the SpaceX REST API and SpaceX wiki page. These data were converted into a Pandas library where they were processed for further analysis. The details of the data collection method are shown in the next section.

Data Collection – SpaceX API

1. Import Requests, Pandas, Numpy, and Datetime libraries
2. Use Requests to download SpaceX launch history data from REST API

```
requests.get(spacex_url)
```
3. Convert the downloaded JSON file to Pandas library

```
pd.json_normalize(file.json())
```
4. Collect the rocket, payload, launchpad, and the launch outcome data from the appropriate REST API

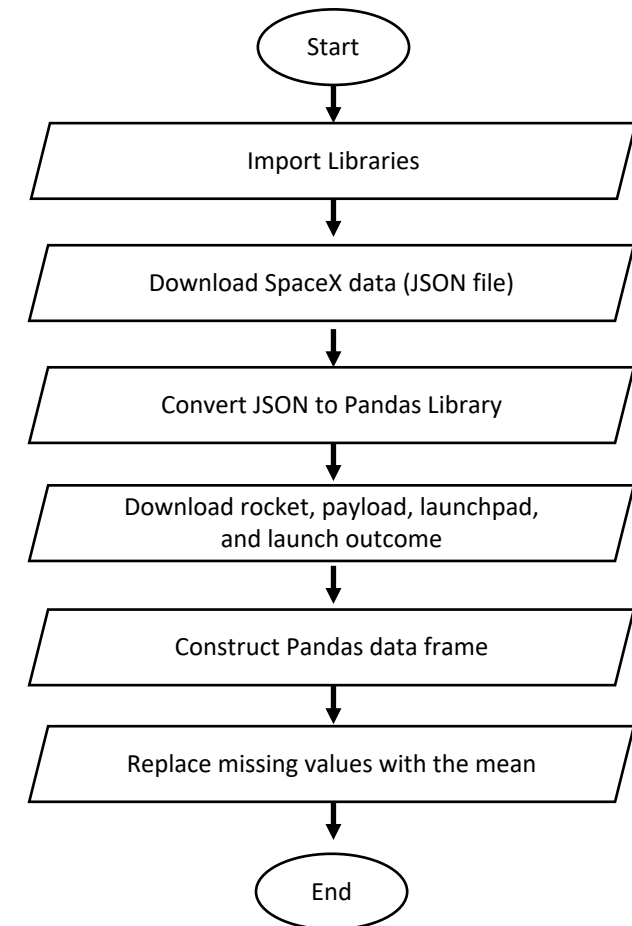
```
getBoosterVersion(data), getLaunchSite(data), getPayloadData(data),  
getCoreData(data)
```
5. Construct Pandas data frame

```
launch_dict = {'FlightNumber': list(data['flight_number']),  
'Date': list(data['date']), 'BoosterVersion':BoosterVersion, ...  
data_falcon = pd.DataFrame.from_dict(launch_dict)
```
6. Check for missing values and replace missing values with the mean

```
data_falcon9.isnull().sum()  
mean = data_falcon9['PayloadMass'].mean()  
data_falcon9['PayloadMass'].fillna(value=mean, inplace=True)
```

The completed SpaceX API calls notebook can be found on this GitHub URL:

[https://github.com/johnisaacenriquez/learning/blob/main/IBM Data Science Capstone Project/Data%20Collection%20API.ipynb](https://github.com/johnisaacenriquez/learning/blob/main/IBM%20Data%20Science%20Capstone%20Project/Data%20Collection%20API.ipynb)



Data Collection - Scraping

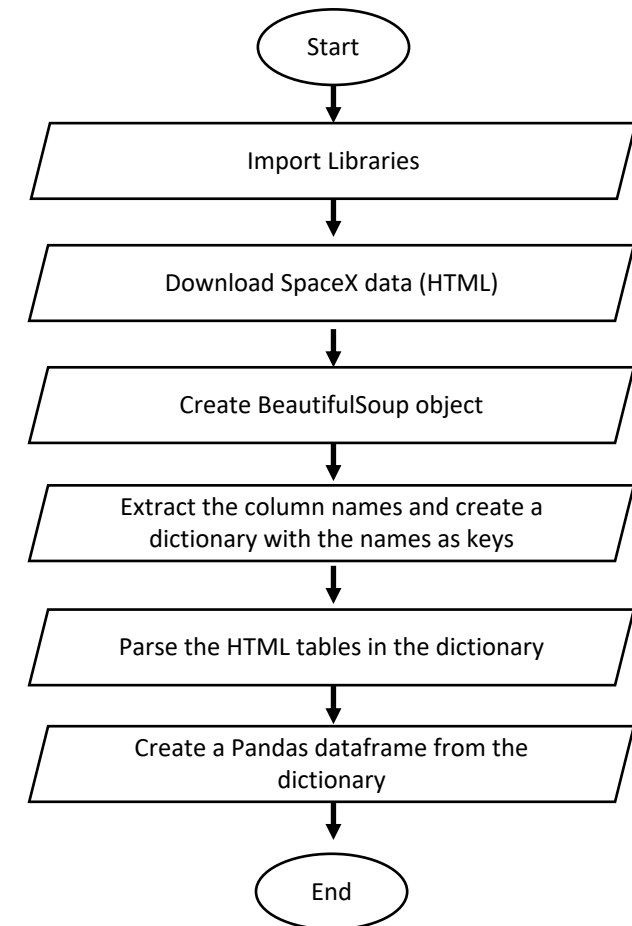
1. Import Requests, BeautifulSoup, and Pandas libraries
2. Use Requests to download and create an object from the Falcon 9 Launch Wiki page

```
page = requests.get(static_url)
```
3. Create a BeautifulSoup object and use find_all function to parse elements with type 'table'

```
soup = BeautifulSoup(page.text, 'html5lib')  
html_tables = soup.find_all('table')  
first_launch_table = html_tables[2]
```
4. Printing the first_launch_table will show that the column names are embedded in table header elements <th>. Select 'th' elements using find_all and extract the columns using the extract_column_from_header() function.

```
column_names = []  
temp = soup.find_all('th')  
for x in range(len(temp)):  
    try:  
        name = extract_column_from_header(temp[x])  
        if (name is not None and len(name) > 0):  
            column_names.append(name)  
    except:  
        pass
```
5. Create an empty dictionary with keys from the extracted columns

```
launch_dict['Flight No.'] = []  
launch_dict['Launch site'] = []  
...
```



Data Collection - Scraping

6. Parse the launch HTML tables to the launch dictionary . Use the provided helper functions if appropriate.

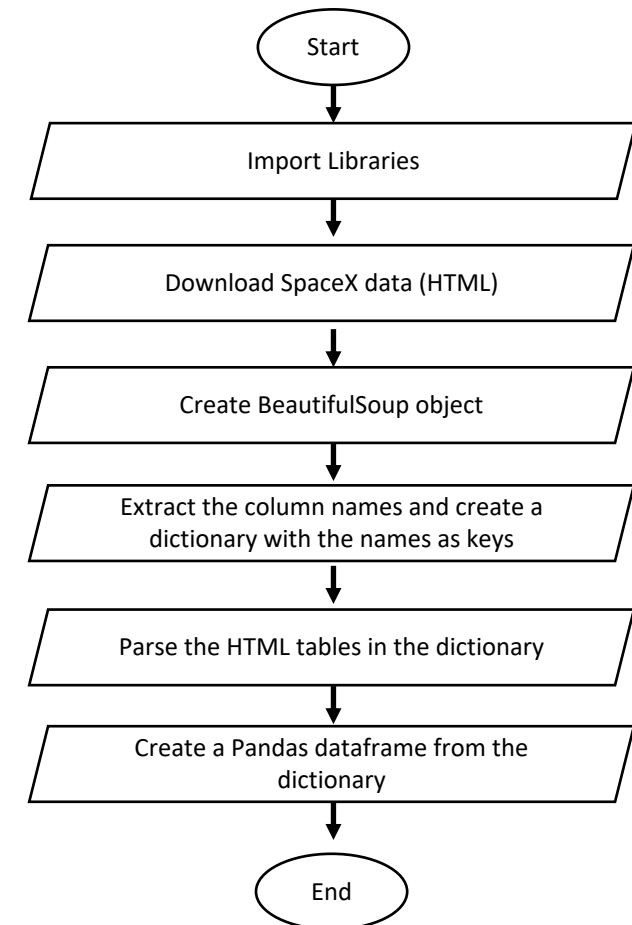
```
for table_number, table in
    enumerate(soup.find_all('table', "wikitable plainrowheaders
        collapsible")):
    for rows in table.find_all("tr"):
        ...
        date = datatimelist[0].strip(',')
        launch_dict["Date"].append(date)
        time = datatimelist[1]
        launch_dict["Time"].append(time)
        launch_site = row[2].a.string
        launch_dict["Launch site"].append(launch_site)
        payload_mass = get_mass(row[4])
        launch_dict["Payload mass"].append(payload_mass)
        ...
```

7. Delete keys in launch dictionary with None values. Finally, create a dataframe from the launch dictionary

```
df = pd.DataFrame.from_dict(launch_dict)
```

The completed web scraping notebook can be found on this GitHub URL:

[https://github.com/johnisaacenriquez/learning/blob/main/IBM Data Science Capstone Project/Data%20Collection%20Web scraping.ipynb](https://github.com/johnisaacenriquez/learning/blob/main/IBM%20Data%20Science%20Capstone%20Project/Data%20Collection%20Web scraping.ipynb)



Data Wrangling

Exploratory data analysis was performed on SpaceX Launch dataframe to find some patterns in the data and determine the label for the landing outcome column.

1. Import Pandas and Numpy libraries

2. Import SpaceX Launch csv file into Pandas

```
df=pd.read_csv("https://cf-courses-  
data.s3.us.cloud-object-  
storage.appdomain.cloud/IBM-DS0321EN-  
SkillsNetwork/datasets/dataset_part_1.csv")
```

3. Identify the amount of missing table values and the types of data

```
df.isnull().sum()/df.count()*100  
df.dtypes
```

4. Calculate the number of launches on each launch site

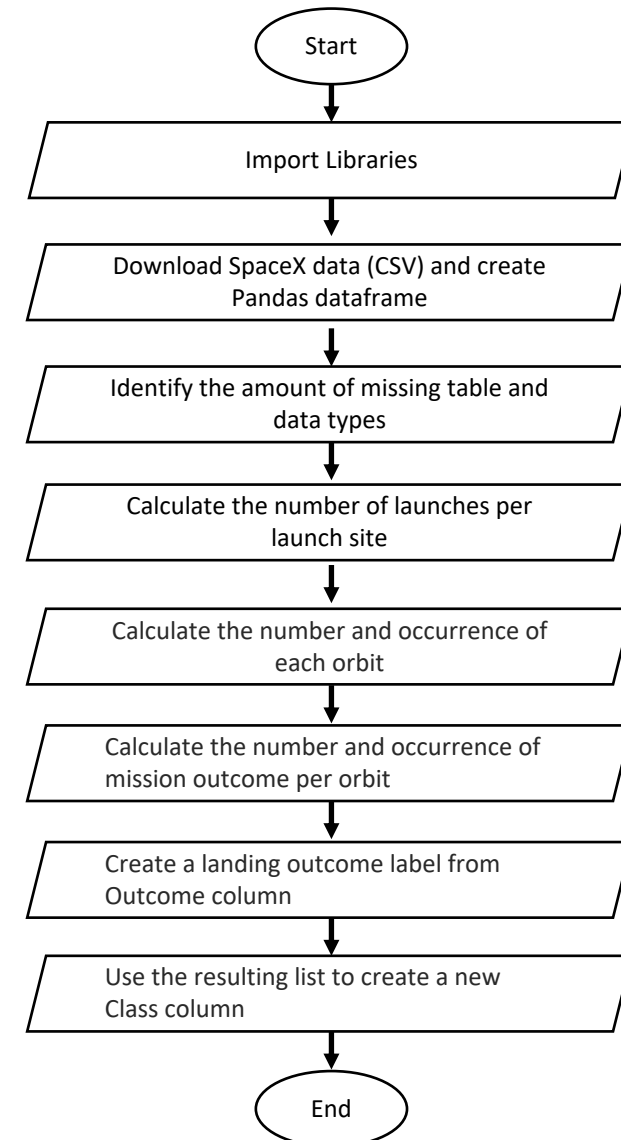
```
df[['LaunchSite']].value_counts()
```

5. Calculate the number and occurrence of each orbit

```
df['Orbit'].value_counts()
```

6. Calculate the number and occurrence of mission outcome per orbit

```
landing_outcomes = df['Outcome'].value_counts()
```



Data Wrangling

7. The 'True' and 'False' values means successful and unsuccessful landings, respectively. Create a list of unsuccessful (bad) outcome types.

```
bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])
```

8. Create a landing outcome label from Outcome column, setting 0 and 1 for unsuccessful and successful outcomes, respectively.

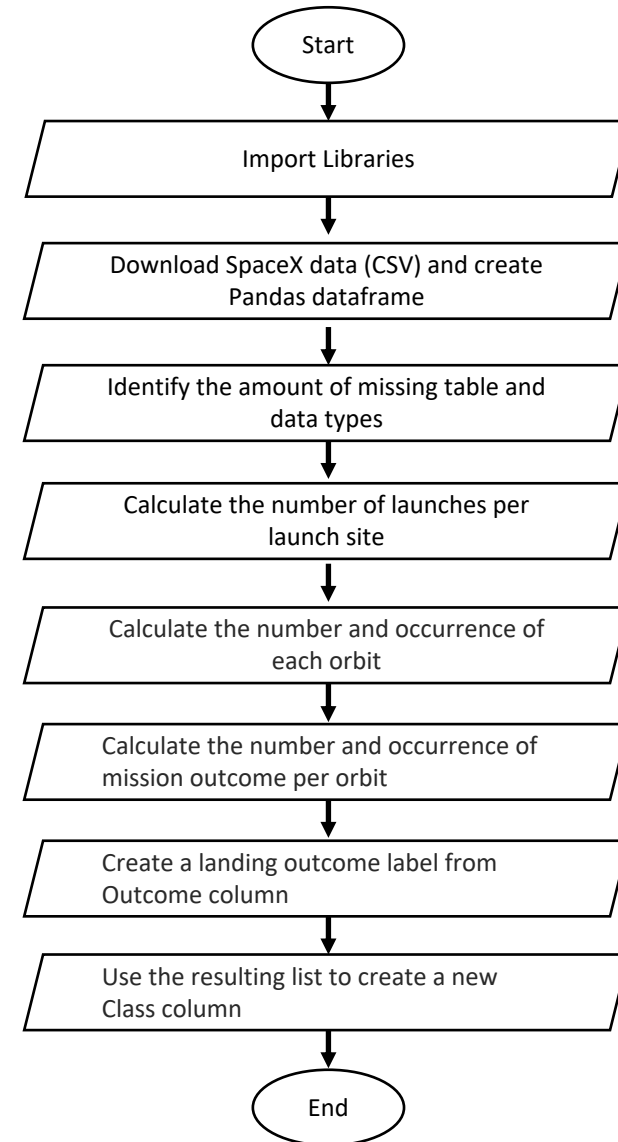
```
landing_class = []
for key,value in df["Outcome"].items():
    if value in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)
```

9. Create a new column 'Class' on the dataframe from the landing_class list.

```
df['Class']=landing_class
```

The completed data wrangling notebook can be found on this GitHub URL:

[https://github.com/johnisaacenriquez/learning/blob/main/IBM Data Science Capstone Project/Data%20Wrangling.ipynb](https://github.com/johnisaacenriquez/learning/blob/main/IBM%20Data%20Science%20Capstone%20Project/Data%20Wrangling.ipynb)



EDA with Data Visualization

Scatterplots and categorical scatterplots were used to investigate the relationship between the following variables:

1. Payload vs Flight Number (catplot)
2. Flight Number vs Launch Site (catplot)
3. Payload vs Launch Site (catplot)
4. Flight Number vs Launch Site (catplot)
5. Flight Number vs Orbit (scatter plot)
6. Payload vs Orbit (scatter plot)

The success rate on each orbit is graphed using bar chart.

Finally, the success rate as a function of time (year) was plotted using a line chart.

The completed EDA with data visualization notebook can be found on this GitHub URL:

[https://github.com/johnisaacenriquez/learning/blob/main/IBM Data Science Capstone Project/EDA%20with%20Data%20Visualization.ipynb](https://github.com/johnisaacenriquez/learning/blob/main/IBM%20Data%20Science%20Capstone%20Project/EDA%20with%20Data%20Visualization.ipynb)

Build an Interactive Map with Folium

Folium was used to mark launch sites on a map, showing which of these sites has successful and unsuccessful landings.

- Markers and circles were created to mark the launch sites
- Cluster of markers were created to mark successful launches (green) and unsuccessful launches (red)
- The distances between a launch site and nearby railway, highway, and city were calculated. The distances were represented by lines and labeled using markers

The completed Interactive Visual Analytics with Folium notebook can be found on this GitHub URL:

[https://github.com/johnisaacenriquez/learning/blob/main/IBM Data Science Capstone Project/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb](https://github.com/johnisaacenriquez/learning/blob/main/IBM%20Data%20Science%20Capstone%20Project/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb)

Build a Dashboard with Plotly Dash

Interactive visual analytics is performed by building a Plotly Dash Application. A pie chart that shows the success rate of each launch site, and the relative success rate between each launch site is added. The type of pie chart is selected using a drop down object. In addition, a scatter chart that shows the success of each booster version for every launch site is added. A slider that controls the payload range for the booster versions is also added.

The completed Plotly Dash lab Python code can be found on this GitHub URL:

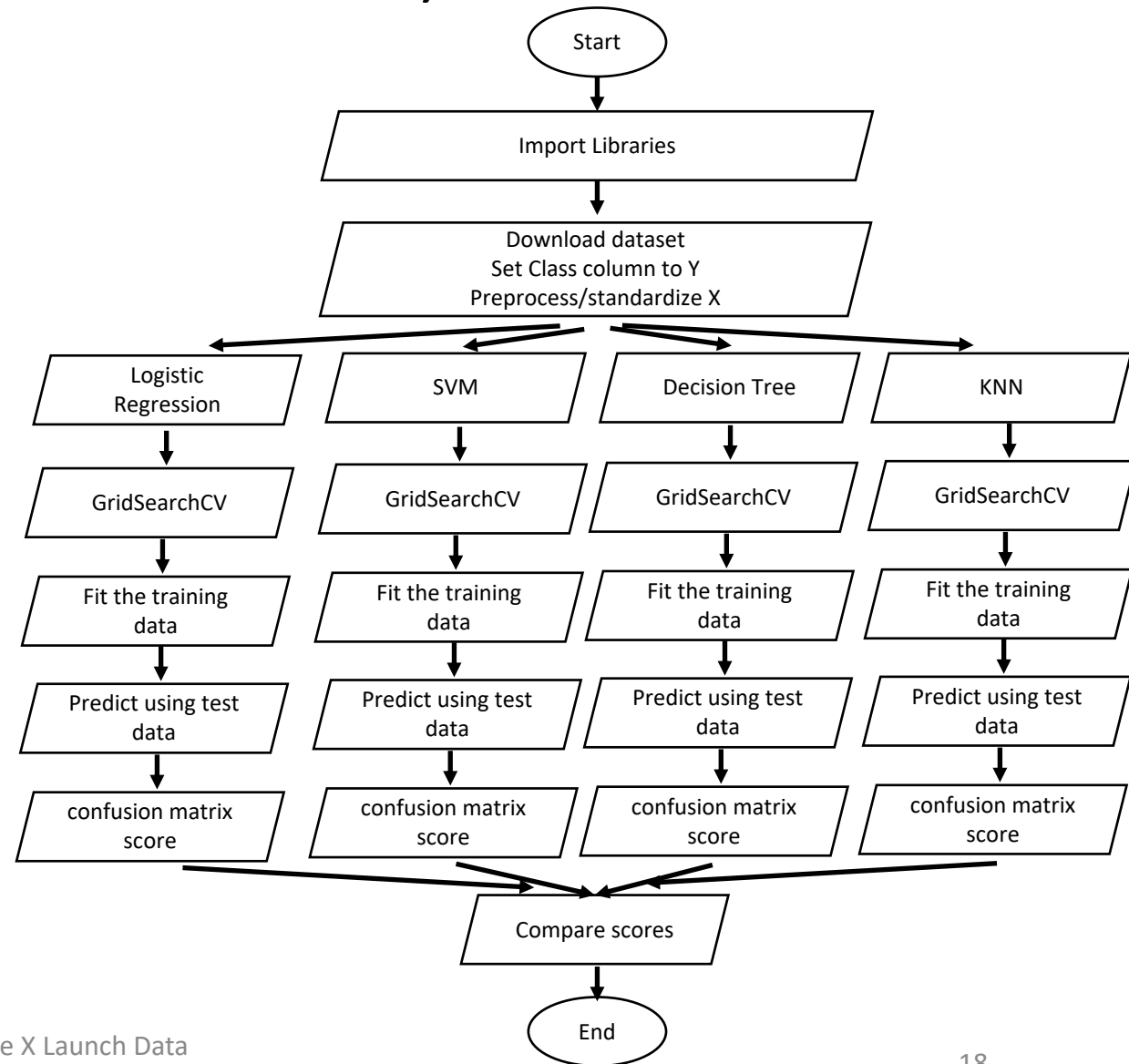
[https://github.com/johnisaacenriquez/learning/blob/main/IBM Data Science Capstone Project/spacex_dash_app.py](https://github.com/johnisaacenriquez/learning/blob/main/IBM%20Data%20Science%20Capstone%20Project/spacex_dash_app.py)

Predictive Analysis (Classification)

Logistic regression, support vector machine, decision tree, and K nearest neighbor classification models were built from scikit-learn library. The features were preprocessed using StandardScaler. The data were split into training and testing data sets. For each classification model, GridSearchCV was used to search over and choose the best parameter values. Each parameter were used to fit the training data. The models were then used to make predictions using the test data. The results of the prediction were presented using confusion matrix. The accuracy of each model was calculated using Score function.

The completed Plotly Dash lab Python code can be found on this GitHub URL:

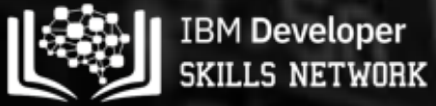
[https://github.com/johnisaacenriquez/learning/blob/main/IBM Data Science Capstone Project/SpaceX Machine%20Learning%20Prediction Part 5.ipynb](https://github.com/johnisaacenriquez/learning/blob/main/IBM%20Data%20Science%20Capstone%20Project/SpaceX%20Machine%20Learning%20Prediction%20Part%205.ipynb)



Section 4

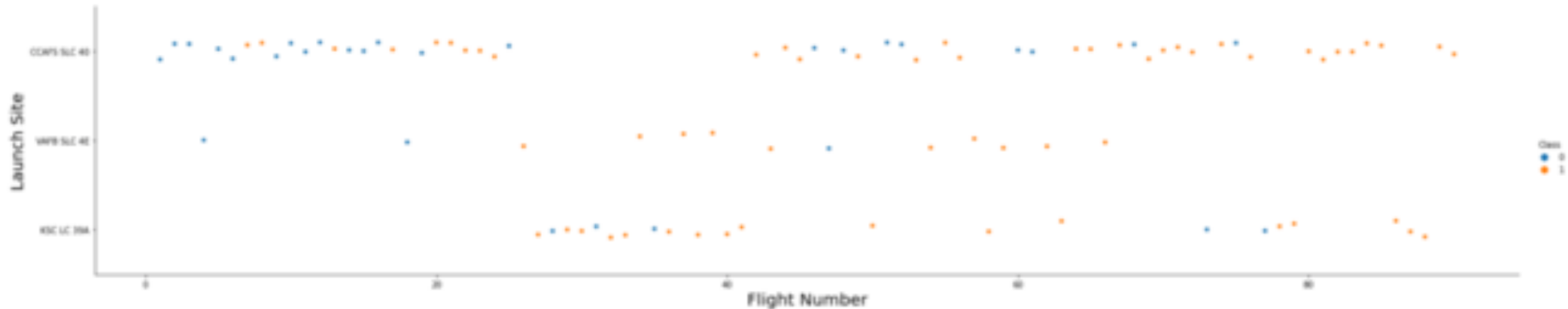
Results





EDA with Visualization

1. Flight Number vs. Launch Site



The figure shows the scatter plot of the flight number vs launch site. Successful and unsuccessful landing of the first stage rocket were represented by blue and orange data points, respectively.

At earlier SpaceX operations (Flight number < 20), the rockets were launched mostly from CCAFS SLC launch site and the success rate were fairly mixed. Two rockets were launched from VAFB SLC but both were unsuccessful. At succeeding SpaceX operations (Flight number > 20), the launches from VAFB SLC and KSC LC were mostly successful. While the launches at CCAFS SLC have mixed result, it has been all successful for the most recent SpaceX operations (Flight number < 80).

In general, most of the launches above Fight number 60 have been successful for all launch sites.

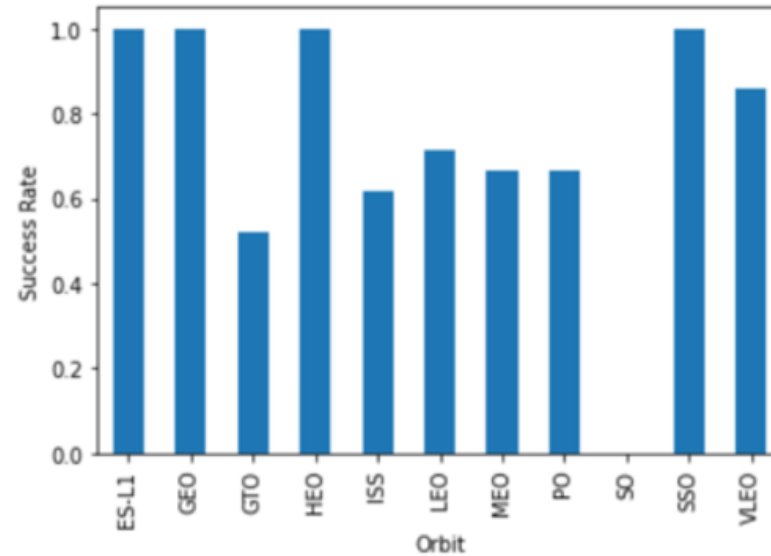
2. Payload vs. Launch Site



The figure shows the scatter plot of the flight number vs pay load mass. Successful and unsuccessful landing of the first stage rocket were represented by blue and orange data points, respectively.

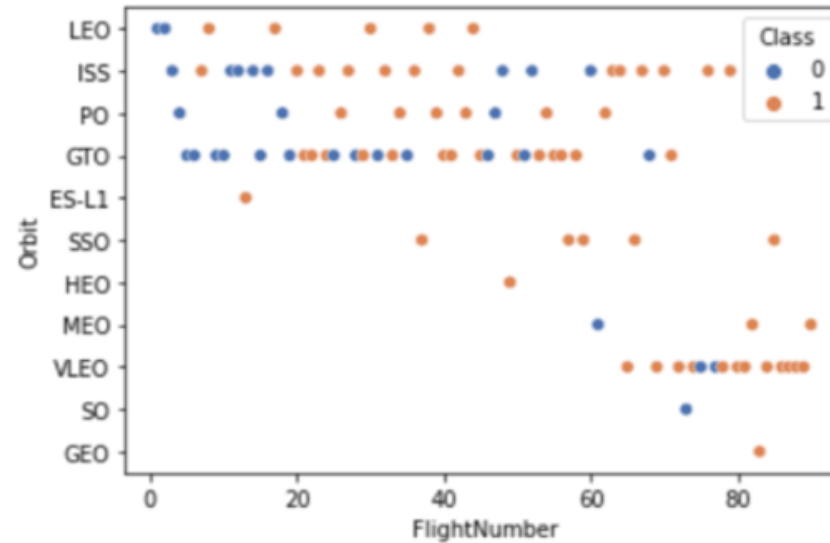
Most of the rockets with payloads of exactly 10000 kg were launched from VAFB SLC. Payloads heavier than 10000 kg were launched from either CCAFS SLC or KSK LC. While most of the rockets were launched from CCAFS SLC, rockets with payloads of between ~8000 kg to ~13000 kg were not launched from this site.

3. Success Rate vs. Orbit Type



The bar chart shows the success rate of recovering the first stage rocket for each orbit. Rocket launched to the ES-L1, GEO, HEO, SSO, and VLEO orbits have been all successful. LEO, MEO, PO, ISS, and GTO have success rates of between 0.5-0.7. A rocket launched to SO has resulted to failure.

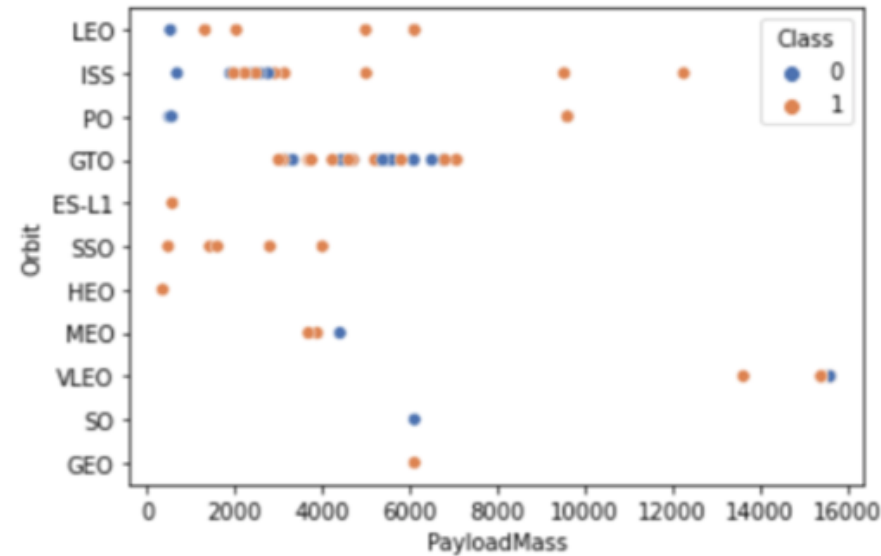
4. Flight Number vs. Orbit Type



The figure shows the scatter plot of the flight number vs orbit type. Successful and unsuccessful landing of the first stage rocket were represented by blue and orange data points, respectively.

Most rockets have been launched to ISS and GTO orbit. Rockets have been launched to ISS orbit from earlier flights to the most recent flights. The frequency of rocket launch to GTO has decreased recently. Since flight number 60, significant number of launches were aimed to VLEO orbit, most of which are successful.

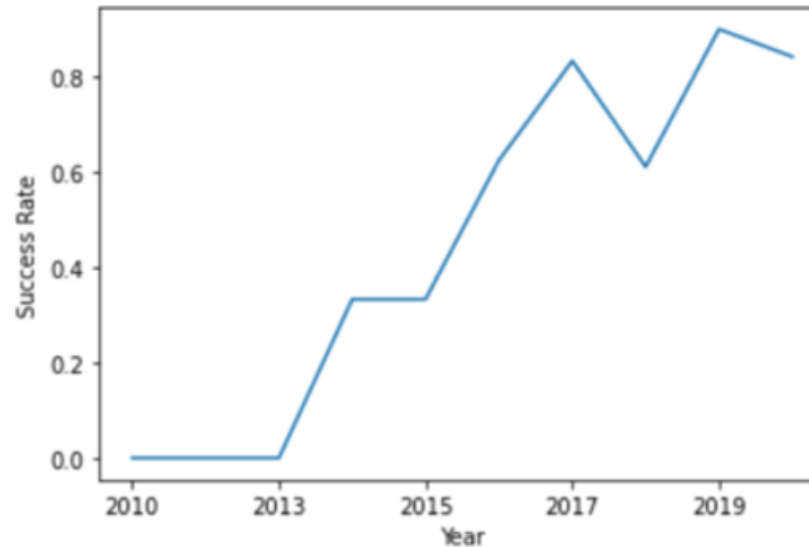
5. Payload vs. Orbit Type



The figure shows the scatter plot of the payload mass vs orbit type. Successful and unsuccessful landing of the first stage rocket were represented by blue and orange data points, respectively.

Heavy payloads (10000 kg or more) were launched mostly to ISS and VLEO orbits. The payloads launched to ISS range from the lightest to the heaviest, while only heavy payloads were launched to VLEO. The payload mass of rocket launched to GTO is convoluted to values between 3000 – 8000 kg.

6. Launch Success Yearly Trend



The figure shows the yearly trend of the launch success. From 2010-2013, all attempts to land the first stage rocket results to failure. Successful launches were recorded between 2013-2014. The success rate had been stagnant from 2014 to 2015 (~0.3), but increased significantly from 2015 to 2017 (0.8). A decrease in the success rate was recorded in 2018 (0.6) but it jumps back up to around 0.9 in 2019.



EDA with SQL

```
BETWEEN 50001 and 1000000 and a.Anno="
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2688
2689
2690
2691
2692
2693
2694
2695
2696
2697
2698
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2729
2730
2731
2732
2733
2734

```

1. All Launch Site Names

Unique launch site names were queried from the data base using the command

```
%sql SELECT  
  UNIQUE (launch_site) FROM  
  SPACEXTBL
```

where %sql is the magic function for Jupyter notebook, SPACEXTBL is the name of the table, and LAUNCH_SITE is the column header.

launch_site

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

2. Launch Site Names Begin with 'CCA'

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Five records where launch sites begin with 'CCA' were queried using the command

```
%sql SELECT * FROM SPACEXTBL WHERE (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5
```

where %sql is the magic function for Jupyter notebook, SPACEXTBLE is the name of the table, and LAUNCH_SITE is the column header. The '%' character in 'CCA%' is a wildcard character.

3. Total Payload Mass

The total payload carried by boosters from NASA were queried from the data base using the command

```
%sql      SELECT      SUM(payload_mass__kg_)      AS  
sum_payload FROM SPACEXTBL WHERE (customer) =  
'NASA (CRS) '
```

sum_payload

45596

where %sql is the magic function for Jupyter notebook, SPACEXTBL is the name of the table, and PAYLOAD_MASS__KG is the column header of the payload mass column.

4. Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 were queried from the data base using the command

```
%sql SELECT AVG(payload_mass__kg_) AS  
average_payload FROM SPACEXTBL WHERE  
(booster_version) = 'F9 v1.1'
```

average_payload

2928

where %sql is the magic function for Jupyter notebook, SPACEXTBL is the name of the table, PAYLOAD_MASS__KG is the column header of the payload mass column, and BOOSTER_VERSION is the column header for the name of boosters.

5. First Successful Ground Landing Date

The date of the first successful landing outcome on ground pad were queried from the data base using the command

```
%sql SELECT MIN(date) FROM SPACEXTBL WHERE  
landing__outcome = 'Success (ground pad) '
```

1

2015-12-22

where %sql is the magic function for Jupyter notebook, SPACEXTBLE is the name of the table, and LANDING__OUTCOME is the column header for the landing outcomes.

6. Successful Drone Ship Landing with Payload between 4000 and 6000 kg

The names of boosters which have successfully landed on drone ship and had payload greater than 4000 kg but less than 6000 kg were queried from the data base using the command

```
%sql SELECT booster_version FROM SPACEXTBL  
WHERE landing__outcome = 'Success (drone  
ship)' AND payload_mass__kg_ BETWEEN 4000  
AND 6000
```

where %sql is the magic function for Jupyter notebook, SPACEXTBL is the name of the table, LANDING__OUTCOME is the column header for the landing outcomes, and PAYLOAD_MASS__KG_ is the column header for the payload mass.

booster_version

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

7. Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes were queried from the data base using the command

```
%sql      SELECT      mission_outcome,  
COUNT(mission_outcome) AS outcome FROM  
SPACEXTBL GROUP BY mission_outcome
```

where %sql is the magic function for Jupyter notebook, SPACEXTBL is the name of the table, MISSION_OUTCOME is the column header for the mission outcomes

mission_outcome	outcome
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

8. Boosters that Carried Maximum Payload

The names of the booster which have carried the maximum payload mass were queried from the data base using the command

```
%sql SELECT booster_version FROM
SPACEXTBL WHERE payload_mass__kg_ =
(SELECT MAX(payload_mass__kg_) FROM
SPACEXTBL)
```

where %sql is the magic function for Jupyter notebook, SPACEXTBL is the name of the table, BOOSTER_VERSION is the column header for the names of the boosters, and PAYLOAD_MASS__KG is the column header for the the mass of the payloads

booster_version

F9 B5 B1048.4

F9 B5 B1049.4

F9 B5 B1051.3

F9 B5 B1056.4

F9 B5 B1048.5

F9 B5 B1051.4

F9 B5 B1049.5

F9 B5 B1060.2

F9 B5 B1058.3

F9 B5 B1051.6

F9 B5 B1060.3

F9 B5 B1049.7

9. 2015 Launch Records

The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 were queried from the data base using the command

```
%sql          SELECT          date,
booster_version,    launch_site,
landing__outcome FROM SPACEXTBL
WHERE      landing__outcome      =
'Failure      (drone      ship)' AND
YEAR(date) = '2015'
```

where %sql is the magic function for Jupyter notebook, SPACEXTBLE is the name of the table, LAUNCH_SITE, LANDING__OUTCOME, and YEAR(date) are the column headers for the launch sites, landing outcomes, and year of the launch date.

DATE	booster_version	launch_site	landing__outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

10. Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20 were queried from the data base and ranked using the command

```
%sql SELECT landing__outcome, COUNT(*)  
AS count_launches FROM SPACEXTBL WHERE  
DATE BETWEEN '2010-06-04' AND '2017-03-  
20' GROUP BY landing__outcome ORDER BY  
count_launches DESC
```

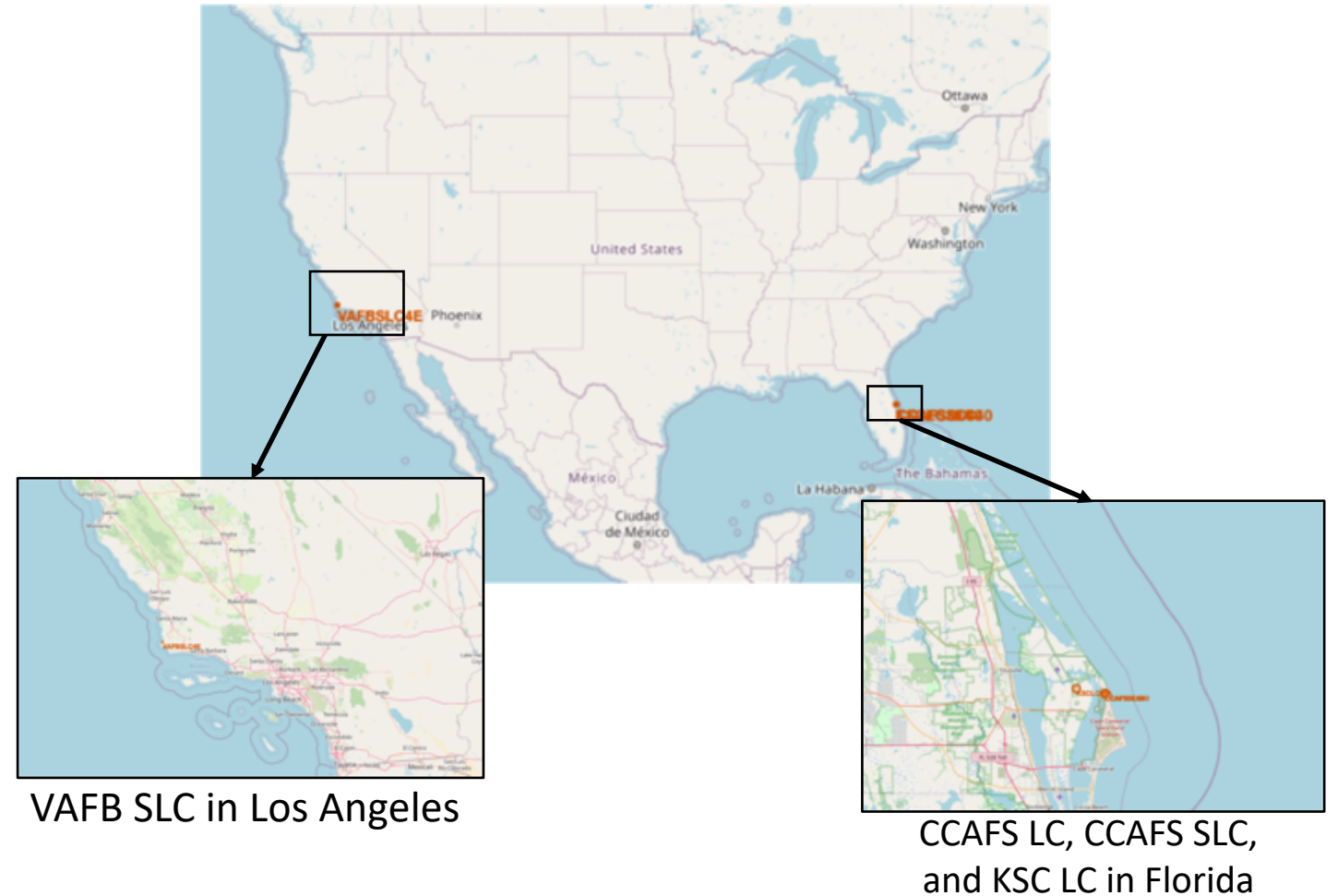
where %sql is the magic function for Jupyter notebook, SPACEXTBL is the name of the table, LANDING__OUTCOME is the column header for the landing outcomes, and COUNT_LAUNCHES is the column header for representing total number of launches calculated from the total number of rows

landing__outcome	count_launches
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Launch Sites Proximities Analysis

1. Launch Site Locations

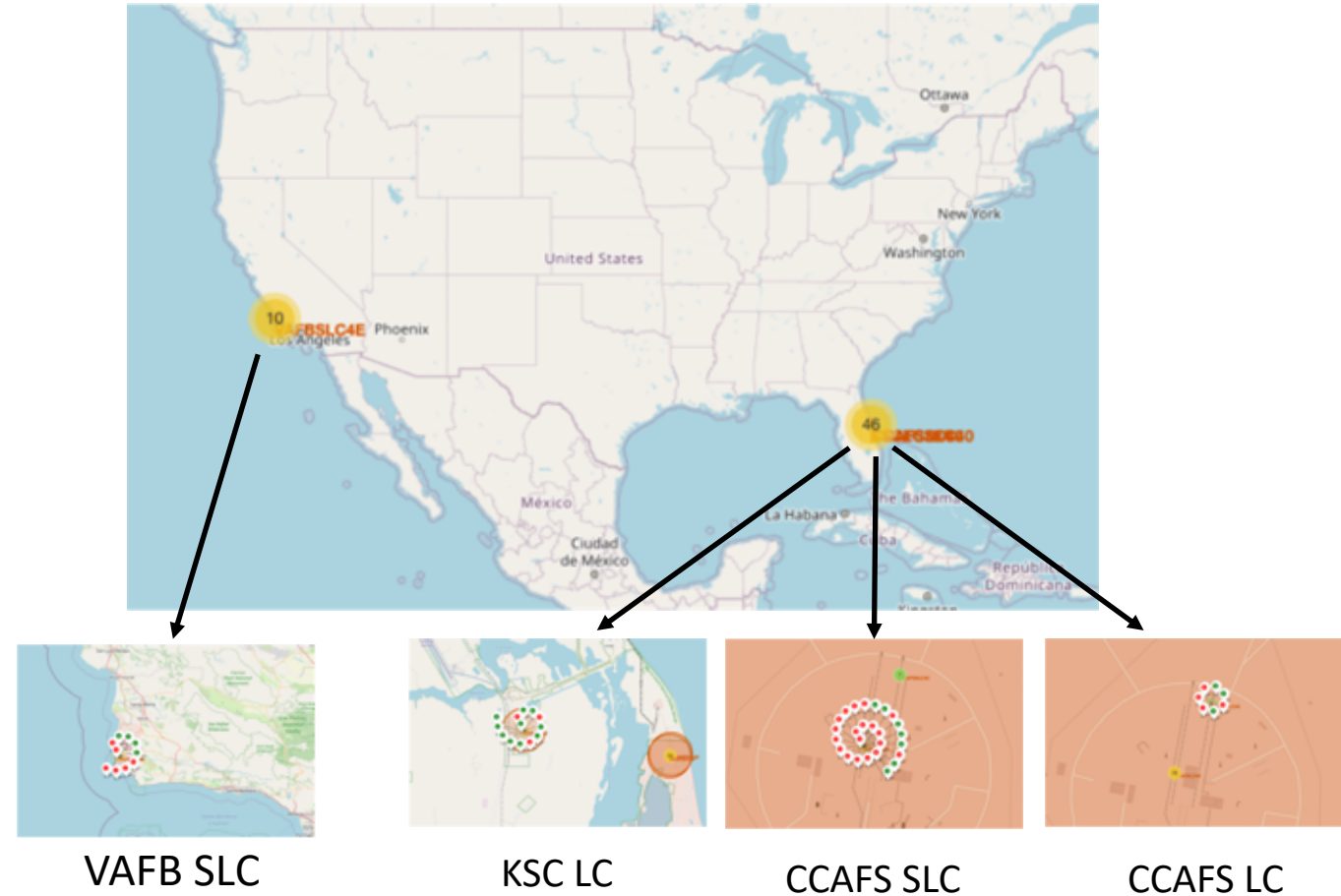
Launch sites were located either in Los Angeles or Florida. All launch sites are in close proximity to the coast and are located in the southernmost part of the United States.



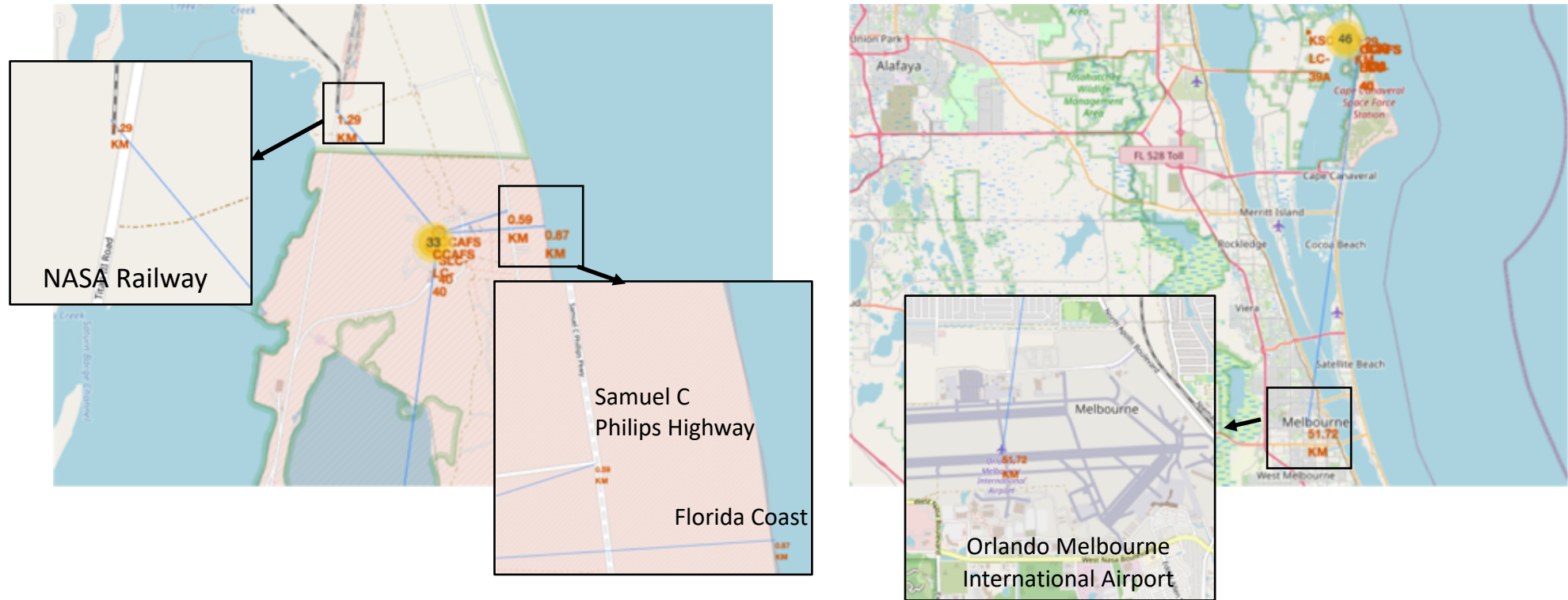
2. Launch Outcomes

The map shows the number of launches in each area. A total of 10 launches were done on VAFB SLC site in Los Angeles, while 46 were done on Florida, of which 13, 26, and 7 were from KSC LC, CCAFS SLC, and CCAFS LC launch sites, respectively.

The green and red markers on the map show the successful and failed launch outcomes. KSC LC has the best success-fail ratio of 10:3. It is followed by CCAFS LC, VAFB SLC, and CCAFS SLC with success-fail ratios of 3:4, 4:6, 7:19, respectively.

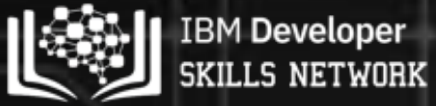


3. Proximity of CCAFS SLC Launch Site to Common Landmarks



The map shows the proximity of CCAFS SLC to some common landmarks. The figure on the left shows its proximity to the nearest railway (left inset) and highway (right inset). The distance to the coastline is also shown (right inset).

The figure on the right shows its proximity to the city of Melbourne. The inset shows that the proximity is measured to the Orlando Melbourne International Airport.

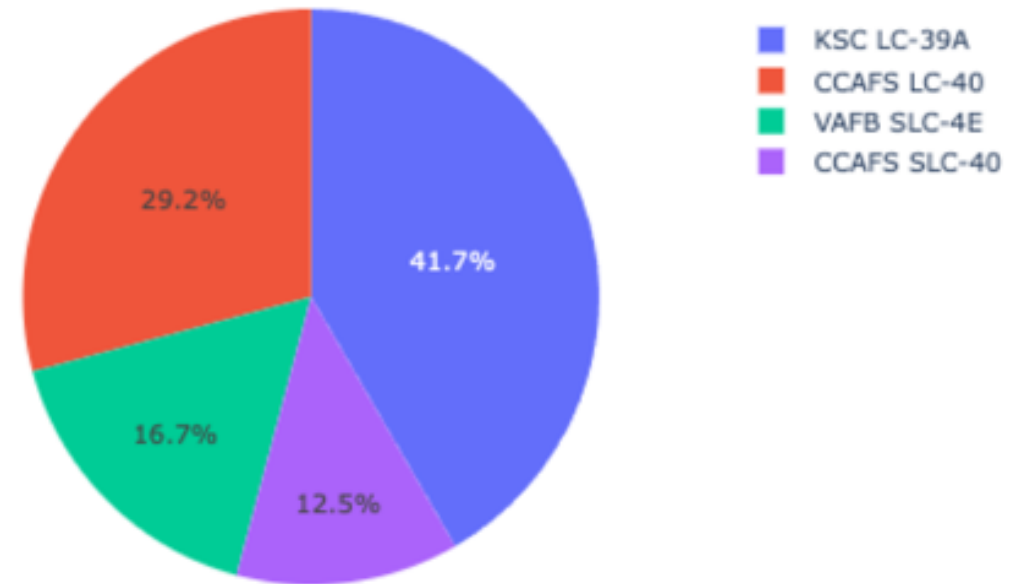


Dashboard with Plotly Dash

1. Launch Success Count for All Launch Sites

The pie graph shows the success count for all launch sites. KSC LC has the most success count, followed by CCAFS LC, VAFB SLC, and CCAFS SLC.

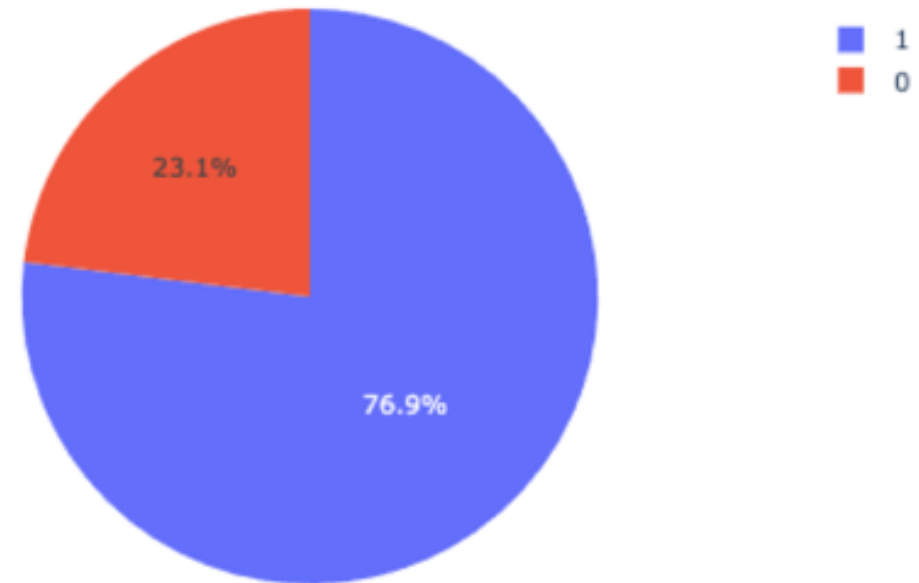
Success Count for all launch sites



2. Total Success Launches for Site KSC LC-39A

The pie graph shows the success rate for the site KSC LC, where 0 and 1 corresponds to failure and success, respectively.

Total Success Launches for site KSC LC-39A



3. Success Count on Payload Mass for All Sites

The figure shows the success count for a given payload mass.

For payloads of less than 4000 kg, five booster versions were used, of which FT booster has the most number of successful outcomes.

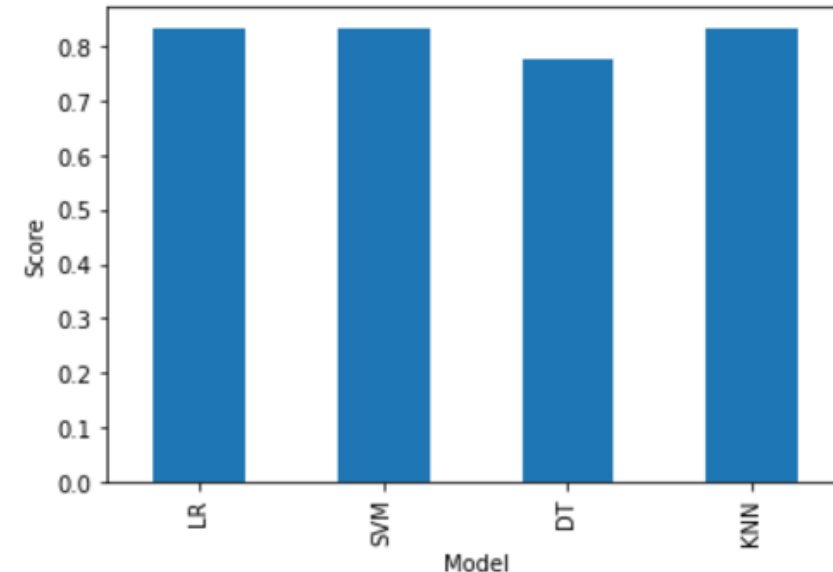
For payloads more than 4000 kg, only three booster versions were used, and only FT and B4 boosters have recorded a successful outcome. For heavy payloads (> 9000 kg), only the B4 booster were used.



Predictive Analysis (Classification)

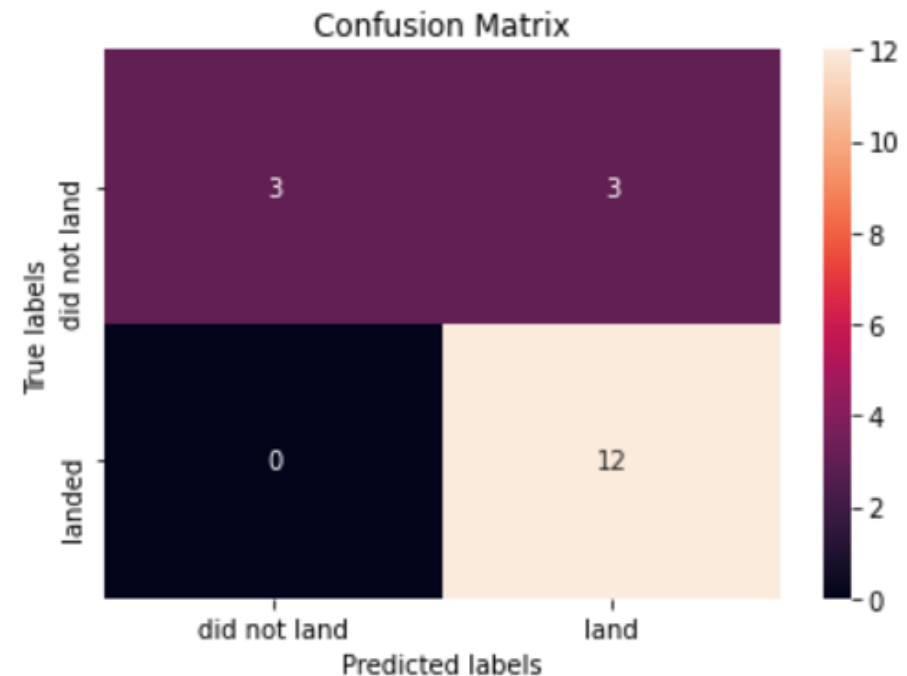
1. Classification Accuracy

The bar chart shows the accuracy of logistic regression (LR), support vector machine (SVM), decision tree (DT), and K nearest neighbor (KNN) classification models calculated using the score function. LR, SVM, KNN are equally accurate in predicting the test data set (0.833), while DT is slightly less accurate (0.778).



2. Confusion Matrix

The confusion matrix for the KNN classifier is shown. There are 3 true negatives (did not land) and 12 true positive (land) predictions. Moreover, there are 0 false negative and 3 false positive. The KNN was able to distinguish between different classes. However, it has problems in giving false positive predictions.



Section 5

Conclusions

Conclusions

- SpaceX launch data were collected and analyzed to generate insights and build machine learning classification model that will predict the success of landing the first stage booster.
- All launch sites are in close proximity to the coasts, and located at the southernmost parts of United states.
- Launch date, launch site, payload weight, and orbit type are the features that correlate to the success of launch.
- In general, the success rate has been increasing every year.
- Machine learning classification models were built to predict the launch outcome. An accuracy score of 0.833 were obtained using logistic regression, support vector machine, and K nearest neighbor.

Appendix

Web scraping helper functions

```
def date_time(table_cells):
    """
    This function returns the data and time from the HTML table cell
    Input: the element of a table data cell extracts extra row
    """
    return [data_time.strip() for data_time in list(table_cells.strings)][0:2]

def booster_version(table_cells):
    """
    This function returns the booster version from the HTML table cell
    Input: the element of a table data cell extracts extra row
    """
    out=''
    for i,booster_version in enumerate(table_cells.strings) if i%2==0 else 0:-1:
        out+=booster_version
    return out

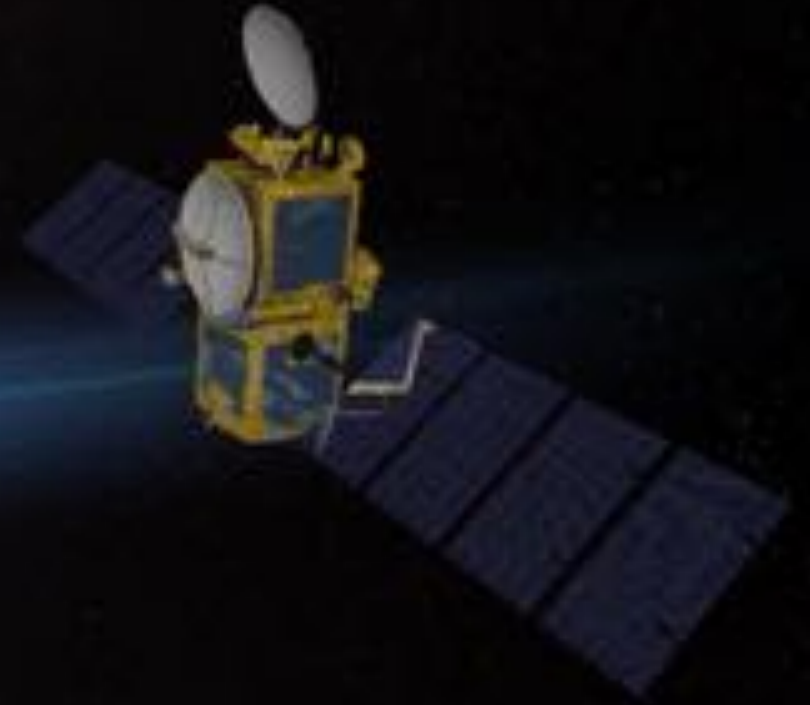
def landing_status(table_cells):
    """
    This function returns the landing status from the HTML table cell
    Input: the element of a table data cell extracts extra row
    """
    out=[i for i in table_cells.strings][0]
    return out

def get_mass(table_cells):
    mass=unicodedata.normalize("NFKD", table_cells.text).strip()
    if mass:
        mass.find("kg")
        new_mass=mass[0:mass.find("kg")+2]
    else:
        new_mass=0
    return new_mass

def extract_column_from_header(row):
    """
    This function returns the landing status from the HTML table cell
    Input: the element of a table data cell extracts extra row
    """
    if (row.br):
        row.br.extract()
    if row.a:
        row.a.extract()
    if row.sup:
        row.sup.extract()

    column_name = ' '.join(row.contents)

    # Filter the digit and empty names
    if not(column_name.strip().isdigit()):
        column_name = column_name.strip()
        return column_name
```



Thank you for your attention.