Assignment: SQL Notebook for Peer Assignment

Estimated time needed: **60** minutes.

Introduction

Using this Python notebook you will:

1. Understand the Spacex DataSet
2. Load the dataset into the corresponding table in a Db2 database
3. Execute SQL queries to answer assignment questions

Overview of the DataSet

SpaceX has gained worldwide attention for a series of historic milestones.

It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars wheras other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.

Therefore if we can determine if the first stage will land, we can determine the cost of a launch.

This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

This dataset includes a record for each payload carried during a SpaceX mission into outer space.

Download the datasets

This assignment requires you to load the spacex dataset.

In many cases the dataset to be analyzed is available as a .CSV (comma separated values) file, perhaps on the internet. Click on the link below to download and save the dataset (.CSV file):

Spacex DataSet

In [1]:
```
!pip install sqlalchemy==1.3.9
```

```
Requirement already satisfied: sqlalchemy==1.3.9 in /home/jupyterlab/conda/en
vs/python/lib/python3.7/site-packages (1.3.9)
```

Connect to the database

Let us first load the SQL extension and establish a connection with the database

In [2]:
```
%load_ext sql
```

In [3]:
```
import csv, sqlite3

con = sqlite3.connect("my_data1.db")
cur = con.cursor()
```

In [4]:
```
!pip install -q pandas==1.1.5
```

In [5]:
```
%sql sqlite:///my_data1.db
```

Out[5]:
```
'Connected: @my_data1.db'
```

In [6]:
```
import pandas as pd
df = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdoma
df.to_sql("SPACEXTBL", con, if_exists='replace', index=False,method="multi")
```

```
/home/jupyterlab/conda/envs/python/lib/python3.7/site-packages/pandas/core/ge
neric.py:2615: UserWarning: The spaces in these column names will not be chan
ged. In pandas versions < 0.14, spaces were converted to underscores.
  method=method,
```

Note:This below code is added to remove blank rows from table

In [7]:
```
%sql create table SPACEXTABLE as select * from SPACEXTBL where Date is not r
```

```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) table SPACEXTABLE already exists
[SQL: create table SPACEXTABLE as select * from SPACEXTBL where Date is not n
ull]
(Background on this error at: http://sqlalche.me/e/e3q8)
```

In [8]:
```
%%sql

select * from SPACEXTABLE;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677
2013-09-29	16:00:00	F9 v1.1 B1003	VAFB SLC-4E	CASSIOPE	500
2013-12-03	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170
2014-01-06	22:06:00	F9 v1.1	CCAFS LC-40	Thaicom 6	3325
2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296
2014-07-14	15:15:00	F9 v1.1	CCAFS LC-40	OG2 Mission 1 6 Orbcomm-OG2 satellites	1316
2014-08-05	8:00:00	F9 v1.1	CCAFS LC-40	AsiaSat 8	4535
2014-09-07	5:00:00	F9 v1.1 B1011	CCAFS LC-40	AsiaSat 6	4428
2014-09-21	5:52:00	F9 v1.1 B1010	CCAFS LC-40	SpaceX CRS-4	2216
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395
2015-02-11	23:03:00	F9 v1.1 B1013	CCAFS LC-40	DSCOVR	570
2015-03-02	3:50:00	F9 v1.1 B1014	CCAFS LC-40	ABS-3A Eutelsat 115 West B	4159

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898
2015-04-27	23:03:00	F9 v1.1 B1016	CCAFS LC-40	Turkmen 52 / MonacoSAT	4707
2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034
2016-01-17	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100
2016-06-15	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070
2017-06-03	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708
2017-06-23	19:10:00	F9 FT B1029.2	KSC LC-39A	BulgariaSat-1	3669
2017-06-25	20:25:00	F9 FT B1036.1	VAFB SLC-4E	Iridium NEXT 2	9600
2017-07-05	23:38:00	F9 FT B1037	KSC LC-39A	Intelsat 35e	6761
2017-08-14	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310
2017-08-24	18:51:00	F9 FT B1038.1	VAFB SLC-4E	Formosat-5	475
2017-09-07	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990
2017-10-09	12:37:00	F9 B4 B1041.1	VAFB SLC-4E	Iridium NEXT 3	9600
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200
2017-10-30	19:34:00	F9 B4 B1042.1	KSC LC-39A	Koreasat 5A	3500
2017-12-15	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205
2017-12-23	1:27:00	F9 FT B1036.2	VAFB SLC-4E	Iridium NEXT 4	9600
2018-01-08	1:00:00	F9 B4 B1043.1	CCAFS SLC-40	Zuma	5000
2018-01-31	21:25:00	F9 FT B1032.2	CCAFS SLC-40	GovSat-1 / SES-16	4230
2018-02-22	14:17:00	F9 FT B1038.2	VAFB SLC-4E	Paz Tintin A & B	2150
2018-03-06	5:33:00	F9 B4 B1044	CCAFS SLC-40	Hispasat 30W-6 PODSat	6092
2018-03-	14:14:00	F9 B4 B1041.2	VAFB SLC-4E	Iridium NEXT 5	9600

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
30					
2018-04-02	20:30:00	F9 B4 B1039.2	CCAFS SLC-40	SpaceX CRS-14	2647
2018-04-18	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362
2018-05-11	20:14:00	F9 B5 B1046.1	KSC LC-39A	Bangabandhu-1	3600
2018-05-22	19:47:58	F9 B4 B1043.2	VAFB SLC-4E	Iridium NEXT 6 GRACE-FO 1, 2	6460
2018-06-04	4:45:00	F9 B4 B1040.2	CCAFS SLC-40	SES-12	5384
2018-06-29	9:42:00	F9 B4 B1045.2	CCAFS SLC-40	SpaceX CRS-15	2697
2018-07-22	5:50:00	F9 B5B1047.1	CCAFS SLC-40	Telstar 19V	7075
2018-07-25	11:39:00	F9 B5B1048.1	VAFB SLC-4E	Iridium NEXT-7	9600
2018-08-07	5:18:00	F9 B5 B1046.2	CCAFS SLC-40	Merah Putih	5800
2018-09-10	4:45:00	F9 B5B1049.1	CCAFS SLC-40	Telstar 18V / Apstar-5C	7060
2018-10-08	2:22:00	F9 B5 B1048.2	VAFB SLC-4E	SAOCOM 1A	3000
2018-11-15	20:46:00	F9 B5 B1047.2	KSC LC-39A	Es hail 2	5300
2018-12-03	18:34:05	F9 B5 B1046.3	VAFB SLC-4E	SSO-A	4000
2018-12-05	18:16:00	F9 B5B1050	CCAFS SLC-40	SpaceX CRS-16	2500
2018-12-23	13:51:00	F9 B5B1054	CCAFS SLC-40	GPS III-01	4400
2019-01-11	15:31:00	F9 B5 B1049.2	VAFB SLC-4E	Iridium NEXT-8	9600
2019-02-22	1:45:00	F9 B5 B1048.3	CCAFS SLC-40	Nusantara Satu, Beresheet	4850

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
				Moon lander, S5	
2019-03-02	7:49:00	F9 B5B1051.1	KSC LC-39A	Crew Dragon Demo-1, SpaceX CRS-17	12055
2019-05-04	6:48:00	F9 B5B1056.1	CCAFS SLC-40	SpaceX CRS-17, Starlink v0.9	2495
2019-05-24	2:30:00	F9 B5 B1049.3	CCAFS SLC-40	Starlink v0.9, RADARSAT Constellation	13620
2019-06-12	14:17:00	F9 B5 B1051.2	VAFB SLC-4E	RADARSAT Constellation, SpaceX CRS-18	4200
2019-07-25	22:01:00	F9 B5 B1056.2	CCAFS SLC-40	SpaceX CRS-18, AMOS-17	2268
2019-08-06	23:23:00	F9 B5 B1047.3	CCAFS SLC-40	AMOS-17, Starlink 1 v1.0	6500
2019-11-11	14:56:00	F9 B5 B1048.4	CCAFS SLC-40	Starlink 1 v1.0, SpaceX CRS-19	15600
2019-12-05	17:29:00	F9 B5B1059.1	CCAFS SLC-40	SpaceX CRS-19, JCSat-18 / Kacific 1	2617
2019-12-17	0:10:00	F9 B5 B1056.3	CCAFS SLC-40	JCSat-18 / Kacific 1, Starlink 2 v1.0	6956
2020-01-07	2:33:00	F9 B5 B1049.4	CCAFS SLC-40	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
2020-01-19	15:30:00	F9 B5 B1046.4	KSC LC-39A	Crew Dragon in-flight abort test, Starlink 3 v1.0	12050
2020-01-29	14:07:00	F9 B5 B1051.3	CCAFS SLC-40	Starlink 3 v1.0, Starlink 4 v1.0	15600
2020-02-17	15:05:00	F9 B5 B1056.4	CCAFS SLC-40	Starlink 4 v1.0, SpaceX CRS-20	15600
2020-03-07	4:50:00	F9 B5 B1059.2	CCAFS SLC-40	SpaceX CRS-20, Starlink 5	1977

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
				v1.0	
2020-03-18	12:16:00	F9 B5 B1048.5	KSC LC-39A	Starlink 5 v1.0, Starlink 6 v1.0	15600
2020-04-22	19:30:00	F9 B5 B1051.4	KSC LC-39A	Starlink 6 v1.0, Crew Dragon Demo-2	15600
2020-05-30	19:22:00	F9 B5B1058.1	KSC LC-39A	Crew Dragon Demo-2, Starlink 7 v1.0	12530
2020-06-04	1:25:00	F9 B5 B1049.5	CCAFS SLC-40	Starlink 7 v1.0, Starlink 8 v1.0	15600
2020-06-13	9:21:00	F9 B5 B1059.3	CCAFS SLC-40	Starlink 8 v1.0, SkySats-16, -17, -18, GPS III-03	15410
2020-06-30	20:10:46	F9 B5B1060.1	CCAFS SLC-40	GPS III-03, ANASIS-II	4311
2020-07-20	21:30:00	F9 B5 B1058.2	CCAFS SLC-40	ANASIS-II, Starlink 9 v1.0	5500
2020-08-07	5:12:00	F9 B5 B1051.5	KSC LC-39A	Starlink 9 v1.0, SXRS-1, Starlink 10 v1.0	14932
2020-08-18	14:31:00	F9 B5 B1049.6	CCAFS SLC-40	Starlink 10 v1.0, SkySat-19, -20, -21, SAOCOM 1B	15440
2020-08-30	23:18:00	F9 B5 B1059.4	CCAFS SLC-40	SAOCOM 1B, GNOMES 1, Tyvak-0172	3130
2020-09-03	12:46:14	F9 B5 B1060.2	KSC LC-39A	Starlink 11 v1.0, Starlink 12 v1.0	15600
2020-10-06	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600
2020-10-18	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
2020-10-24	15:31:34	F9 B5 B1060.3	CCAFS SLC-40	Starlink 14 v1.0, GPS III-04	15600
2020-11-05	23:24:23	F9 B5B1062.1	CCAFS SLC-40	GPS III-04 , Crew-1	4311
2020-11-16	0:27:00	F9 B5B1061.1	KSC LC-39A	Crew-1, Sentinel-6 Michael Freilich	12500
2020-11-21	17:17:08	F9 B5B1063.1	VAFB SLC-4E	Sentinel-6 Michael Freilich, Starlink 15 v1.0	1192
2020-11-25	2:13:00	F9 B5 B1049.7	CCAFS SLC-40	Starlink 15 v1.0, SpaceX CRS-21	15600
2020-				SpaceX CRS-	

Tasks

Now write and execute SQL queries to solve the assignment tasks.

Note: If the column names are in mixed case enclose it in double quotes For Example "Landing_Outcome"

Task 1

Display the names of the unique launch sites in the space mission

In [9]:
```
%sql select DISTINCT "Launch_Site" from SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

Out[9]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Task 2

Display 5 records where launch sites begin with the string 'CCA'

In [10]:
```
%sql select * from SPACEXTABLE where "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[10]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	O
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	L
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	L (I
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	L (I
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	L (I
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	L (I

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

In [11]:
```
%%sql

select SUM("PAYLOAD_MASS__KG_")
from SPACEXTABLE
where Customer = "NASA (CRS)";
```

```
 * sqlite:///my_data1.db
Done.
```

Out[11]:

SUM("PAYLOAD_MASS__KG_")
45596

Task 4

Display average payload mass carried by booster version F9 v1.1

In [12]:
```
%%sql

select AVG("PAYLOAD_MASS__KG_")
from SPACEXTABLE
where "Booster_Version" = "F9 v1.1";
```

```
 * sqlite:///my_data1.db
Done.
```

Out[12]:

AVG("PAYLOAD_MASS__KG_")
2928.4

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

In [13]:
```
%%sql

select min(Date)
from SPACEXTABLE
where "Landing_Outcome" LIKE "Success%";
```

```
 * sqlite:///my_data1.db
Done.
```

Out[13]:

min(Date)
2015-12-22

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [14]:
```
%%sql

select count("Booster_Version")
from SPACEXTABLE
where "Landing_Outcome" == "Success (drone ship)"
AND 4000 < "PAYLOAD_MASS__KG_"
AND "PAYLOAD_MASS__KG_"<6000;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[14]:

count("Booster_Version")
4

Task 7

List the total number of successful and failure mission outcomes

In [15]:
```
%%sql

select Mission_Outcome, count("Mission_Outcome")
from SPACEXTABLE
GROUP BY Mission_Outcome;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[15]:

Mission_Outcome	count("Mission_Outcome")
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [16]:
```
%%sql

select BOOSTER_VERSION as boosterversion
from SPACEXTBL
where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

```
 * sqlite:///my_data1.db
Done.
```

Out[16]:

boosterversion
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

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

In [17]:
```
%%sql

select substr(Date, 6,2) as Month, MISSION_OUTCOME,BOOSTER_VERSION,LAUNCH_SI
FROM SPACEXTBL where substr(Date,0,5)='2015';
```

```
 * sqlite:///my_data1.db
Done.
```

Out[17]:

Month	Mission_Outcome	Booster_Version	Launch_Site
01	Success	F9 v1.1 B1012	CCAFS LC-40
02	Success	F9 v1.1 B1013	CCAFS LC-40
03	Success	F9 v1.1 B1014	CCAFS LC-40
04	Success	F9 v1.1 B1015	CCAFS LC-40
04	Success	F9 v1.1 B1016	CCAFS LC-40
06	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
12	Success	F9 FT B1019	CCAFS LC-40

Task 10

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

In [18]:
```
%%sql

SELECT LANDING_OUTCOME
FROM SPACEXTBL
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY Date DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[18]:

Landing_Outcome
No attempt
Success (ground pad)
Success (drone ship)
Success (drone ship)
Success (ground pad)
Failure (drone ship)
Success (drone ship)
Success (drone ship)
Success (drone ship)
Failure (drone ship)
Failure (drone ship)
Success (ground pad)
Precluded (drone ship)
No attempt
Failure (drone ship)
No attempt
Controlled (ocean)
Failure (drone ship)
Uncontrolled (ocean)
No attempt
No attempt
Controlled (ocean)
Controlled (ocean)
No attempt
No attempt
Uncontrolled (ocean)
No attempt
No attempt
No attempt
Failure (parachute)
Failure (parachute)

Reference Links

- Hands-on Lab : String Patterns, Sorting and Grouping
- Hands-on Lab: Built-in functions
- Hands-on Lab : Sub-queries and Nested SELECT Statements
- Hands-on Tutorial: Accessing Databases with SQL magic
- Hands-on Lab: Analyzing a real World Data Set

Author(s)

Lakshmi Holla

Other Contributors

Rav Ahuja

Change log

Date	Version	Changed by	Change Description
2021-07-09	0.2	Lakshmi Holla	Changes made in magic sql
2021-05-20	0.1	Lakshmi Holla	Created Initial Version

In [ ]:
```
!
```