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 [ ]:

```
!pip install sqlalchemy==1.3.9
!pip install ibm_db_sa
!pip install ipython-sql
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: sqlalchemy==1.3.9 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (1.3.9)
[notice] A new release of pip is available: 23.3.2 -> 24.0
[notice] To update, run: C:\Users\DavidVB\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip
Defaulting to user installation because normal site-packages is not writeable
Collecting ibm_db_sa
  Downloading ibm_db_sa-0.4.0-py3-none-any.whl.metadata (5.3 kB)
Requirement already satisfied: sqlalchemy>=0.7.3 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ibm_db_sa) (1.3.9)
Collecting ibm-db>=2.0.0 (from ibm_db_sa)
  Downloading ibm_db-3.2.3-cp312-cp312-win_amd64.whl.metadata (1.4 kB)
Downloading ibm_db_sa-0.4.0-py3-none-any.whl (31 kB)
Downloading ibm_db-3.2.3-cp312-cp312-win_amd64.whl (27.8 MB)
   ---------------------------------------- 0.0/27.8 MB ? eta -:--:--
    --------------------------------------- 0.5/27.8 MB 14.5 MB/s eta 0:00:02
   - -------------------------------------- 1.2/27.8 MB 15.1 MB/s eta 0:00:02
   -- ------------------------------------- 1.9/27.8 MB 15.2 MB/s eta 0:00:02
   --- ------------------------------------ 2.1/27.8 MB 16.6 MB/s eta 0:00:02
   --- ------------------------------------ 2.8/27.8 MB 12.6 MB/s eta 0:00:02
   ----- ---------------------------------- 3.8/27.8 MB 14.1 MB/s eta 0:00:02
   ------ --------------------------------- 4.7/27.8 MB 15.7 MB/s eta 0:00:02
   -------- ------------------------------- 5.6/27.8 MB 15.5 MB/s eta 0:00:02
   --------- ------------------------------ 6.6/27.8 MB 16.7 MB/s eta 0:00:02
   ---------- ----------------------------- 7.3/27.8 MB 16.7 MB/s eta 0:00:02
   ----------- ---------------------------- 8.0/27.8 MB 16.4 MB/s eta 0:00:02
   ----------- ---------------------------- 9.0/27.8 MB 17.0 MB/s eta 0:00:02
   -------------- ------------------------- 10.1/27.8 MB 17.4 MB/s eta 0:00:02
   ---------------- ----------------------- 11.2/27.8 MB 17.7 MB/s eta 0:00:01
   ----------------- ---------------------- 12.5/27.8 MB 21.1 MB/s eta 0:00:01
   ------------------- -------------------- 13.6/27.8 MB 21.8 MB/s eta 0:00:01
   --------------------- ------------------ 14.9/27.8 MB 22.6 MB/s eta 0:00:01
   ----------------------- ---------------- 16.0/27.8 MB 23.4 MB/s eta 0:00:01
   ------------------------ --------------- 17.2/27.8 MB 23.4 MB/s eta 0:00:01
   -------------------------- ------------- 18.5/27.8 MB 25.2 MB/s eta 0:00:01
   ---------------------------- ----------- 19.5/27.8 MB 25.2 MB/s eta 0:00:01
   ----------------------------- ---------- 20.6/27.8 MB 25.2 MB/s eta 0:00:01
   ------------------------------- -------- 21.7/27.8 MB 25.2 MB/s eta 0:00:01
   -------------------------------- ------- 22.8/27.8 MB 25.2 MB/s eta 0:00:01
   ---------------------------------- ----- 24.0/27.8 MB 25.1 MB/s eta 0:00:01
   ------------------------------------ --- 25.2/27.8 MB 25.1 MB/s eta 0:00:01
   ------------------------------------- -- 26.4/27.8 MB 25.2 MB/s eta 0:00:01
   ---------------------------------------  27.7/27.8 MB 25.2 MB/s eta 0:00:01
   ---------------------------------------  27.8/27.8 MB 25.2 MB/s eta 0:00:01
   ---------------------------------------- 27.8/27.8 MB 21.8 MB/s eta 0:00:00
Installing collected packages: ibm-db, ibm_db_sa
Successfully installed ibm-db-3.2.3 ibm_db_sa-0.4.0
[notice] A new release of pip is available: 23.3.2 -> 24.0
[notice] To update, run: C:\Users\DavidVB\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip
Defaulting to user installation because normal site-packages is not writeable
Collecting ipython-sql
  Downloading ipython_sql-0.5.0-py3-none-any.whl.metadata (17 kB)
Collecting prettytable (from ipython-sql)
  Downloading prettytable-3.10.0-py3-none-any.whl.metadata (30 kB)
Requirement already satisfied: ipython in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython-sql) (8.19.0)
Collecting sqlalchemy>=2.0 (from ipython-sql)
  Downloading SQLAlchemy-2.0.31-cp312-cp312-win_amd64.whl.metadata (9.9 kB)
Collecting sqlparse (from ipython-sql)
  Downloading sqlparse-0.5.0-py3-none-any.whl.metadata (3.9 kB)
Requirement already satisfied: six in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython-sql) (1.16.0)
Collecting ipython-genutils (from ipython-sql)
  Downloading ipython_genutils-0.2.0-py2.py3-none-any.whl.metadata (755 bytes)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from sqlalchemy>=2.0->ipytho
n-sql) (4.9.0)
Collecting greenlet!=0.4.17 (from sqlalchemy>=2.0->ipython-sql)
  Downloading greenlet-3.0.3-cp312-cp312-win_amd64.whl.metadata (3.9 kB)
Requirement already satisfied: decorator in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython->ipython-sql) (5.1.1)
Requirement already satisfied: jedi>=0.16 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython->ipython-sql) (0.19.1)
Requirement already satisfied: matplotlib-inline in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython->ipython-sql) (0.1.6)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython->ipython-s
ql) (3.0.43)
Requirement already satisfied: pygments>=2.4.0 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython->ipython-sql) (2.17.2)
Requirement already satisfied: stack-data in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython->ipython-sql) (0.6.3)
Requirement already satisfied: traitlets>=5 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython->ipython-sql) (5.14.0)
Requirement already satisfied: colorama in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from ipython->ipython-sql) (0.4.6)
Requirement already satisfied: wcwidth in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from prettytable->ipython-sql) (0.2.12)
Requirement already satisfied: parso<0.9.0,>=0.8.3 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from jedi>=0.16->ipython->ipython
-sql) (0.8.3)
Requirement already satisfied: executing>=1.2.0 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from stack-data->ipython->ipython-sq
l) (2.0.1)
Requirement already satisfied: asttokens>=2.1.0 in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from stack-data->ipython->ipython-sq
l) (2.4.1)
Requirement already satisfied: pure-eval in c:\users\davidvb\appdata\local\packages\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-packages\python312\site-packages (from stack-data->ipython->ipython-sql) (0.
2.2)
Downloading ipython_sql-0.5.0-py3-none-any.whl (20 kB)
Downloading SQLAlchemy-2.0.31-cp312-cp312-win_amd64.whl (2.1 MB)
   ---------------------------------------- 0.0/2.1 MB ? eta -:--:--
   --------- ------------------------------ 0.5/2.1 MB 10.7 MB/s eta 0:00:01
   ------------------------------- -------- 1.6/2.1 MB 20.7 MB/s eta 0:00:01
   ---------------------------------------- 2.1/2.1 MB 19.0 MB/s eta 0:00:00
Downloading ipython_genutils-0.2.0-py2.py3-none-any.whl (26 kB)
Downloading prettytable-3.10.0-py3-none-any.whl (28 kB)
Downloading sqlparse-0.5.0-py3-none-any.whl (43 kB)
   ---------------------------------------- 0.0/44.0 kB ? eta -:--:--
   ---------------------------------------- 44.0/44.0 kB ? eta 0:00:00
Downloading greenlet-3.0.3-cp312-cp312-win_amd64.whl (293 kB)
   ---------------------------------------- 0.0/293.6 kB ? eta -:--:--
   --------------------------------------- 293.6/293.6 kB 17.7 MB/s eta 0:00:00
Installing collected packages: ipython-genutils, sqlparse, prettytable, greenlet, sqlalchemy, ipython-sql
  Attempting uninstall: sqlalchemy
    Found existing installation: SQLAlchemy 1.3.9
    Uninstalling SQLAlchemy-1.3.9:
      Successfully uninstalled SQLAlchemy-1.3.9
Successfully installed greenlet-3.0.3 ipython-genutils-0.2.0 ipython-sql-0.5.0 prettytable-3.10.0 sqlalchemy-2.0.31 sqlparse-0.5.0
[notice] A new release of pip is available: 23.3.2 -> 24.0
[notice] To update, run: C:\Users\DavidVB\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip
```

Connect to the database

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

In [ ]:

```
%load_ext sql
```

In [ ]:

```
import csv, sqlite3

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

In [ ]:

```
!pip install -q pandas
```

```
[notice] A new release of pip is available: 23.3.2 -> 24.0
[notice] To update, run: C:\Users\DavidVB\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip
```

In [ ]:

```
%sql sqlite:///my_data1.db
```

In [ ]:

```
import pandas as pd
df = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv")
df.to_sql("SPACEXTBL", con, if_exists='replace', index=False,method="multi")
```

Out[ ]:

```
101
```

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

In [ ]:

```
%sql create table SPACEXTABLE as select * from SPACEXTBL where Date is not null
```

```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) table SPACEXTABLE already exists
[SQL: create table SPACEXTABLE as select * from SPACEXTBL where Date is not null]
(Background on this error at: https://sqlalche.me/e/20/e3q8)
```

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 [ ]:

```
%sql select distinct (Launch_Site) from SPACEXTBL
```

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

Out[ ]:

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 [ ]:

```
%sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
```

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

Out[ ]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

Task 3

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

In [ ]:

```
%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = 'NASA (CRS)'
```

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

Out[ ]:

SUM(PAYLOAD_MASS__KG_)
45596

Task 4

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

In [ ]:

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version LIKE 'F9 v1.0%'
```

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

Out[ ]:

AVG(PAYLOAD_MASS__KG_)
340.4

Task 5

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

Hint:Use min function

In [ ]:

```
%sql select * from SPACEXTBL where Landing_Outcome = 'Success (ground pad)' order by Date asc limit 1
```

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

Out[ ]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

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 [ ]:

```
# %sql select * from SPACEXTBL
%sql select * from SPACEXTBL where Landing_Outcome = 'Success (drone ship)'  and PAYLOAD_MASS__KG_ between 4000 and 6000
```

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

Out[ ]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Task 7

List the total number of successful and failure mission outcomes

In [ ]:

```
%sql select count(*) from SPACEXTBL where Mission_Outcome like 'Success%'

%sql select count(*) from SPACEXTBL where Mission_Outcome like 'Failure%'
```

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

Out[ ]:

count(*)
100

Task 8

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

In [ ]:

```
%sql select distinct (Booster_Version) , PAYLOAD_MASS__KG_ from SPACEXTBL where PAYLOAD_MASS__KG_ in (select PAYLOAD_MASS__KG_ from SPACEXTBL order by PAYLOAD_MASS__KG_ desc limit 12) order by PAYLOAD_MASS__KG_ desc
```

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

Out[ ]:

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

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 [ ]:

```
# %sql select * from SPACEXTBL order by PAYLOAD_MASS__KG_ desc

%sql select * from SPACEXTBL where Landing_Outcome like '%(drone ship)%' and Date between '2015-01-01' and '2015-12-31'
```

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

Out[ ]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)

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 [ ]:

```
%sql select * from SPACEXTBL where Landing_Outcome in ('Failure (drone ship)', 'Success (ground pad)') and Date between '2010-06-04' and '2017-03-20'
```

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

Out[ ]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
2016-01-17	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (drone ship)
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
2016-06-15	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600	GTO	ABS Eutelsat	Success	Failure (drone ship)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)

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