
# CSE 488 (Section 1)
# [Summer 2022]

# Lab 05 Assignment Submission Report

# Assignment Title: Spark Programming

**Submitted by:**
**Rizvee Hassan Prito**
**2019-3-60-041**

# Screenshots:

## Problem 1:

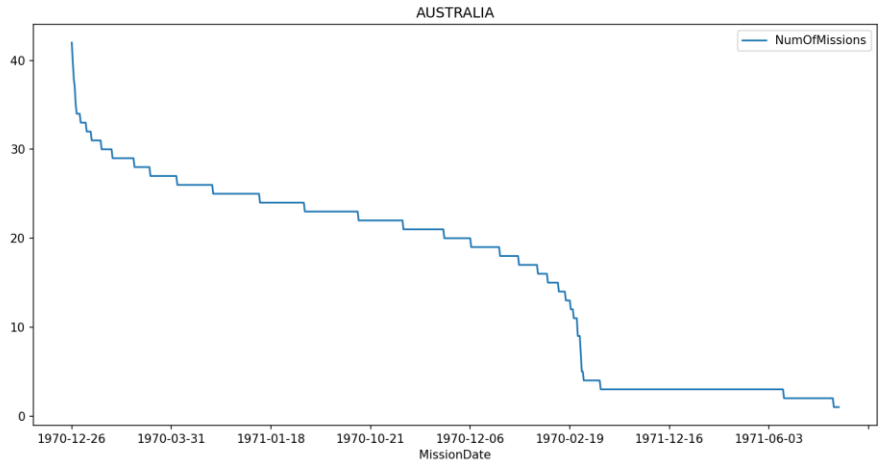
**Find the number of missions completed on each day for each of the countries involved.**

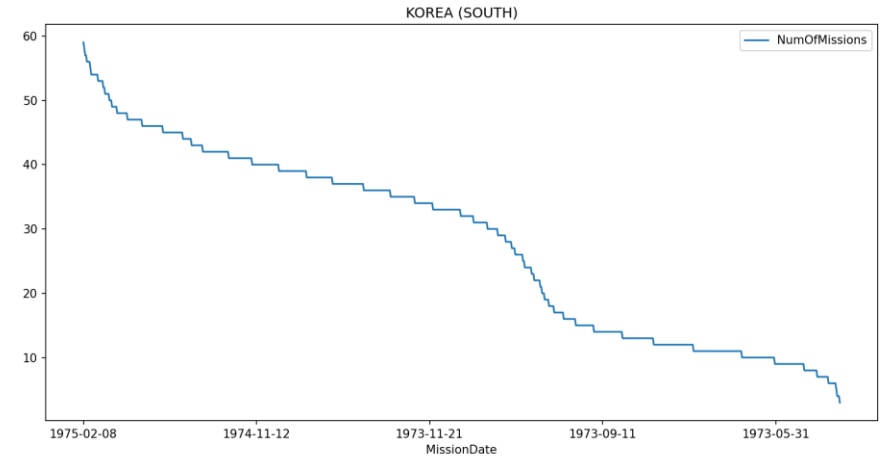
## Codes and Outputs:

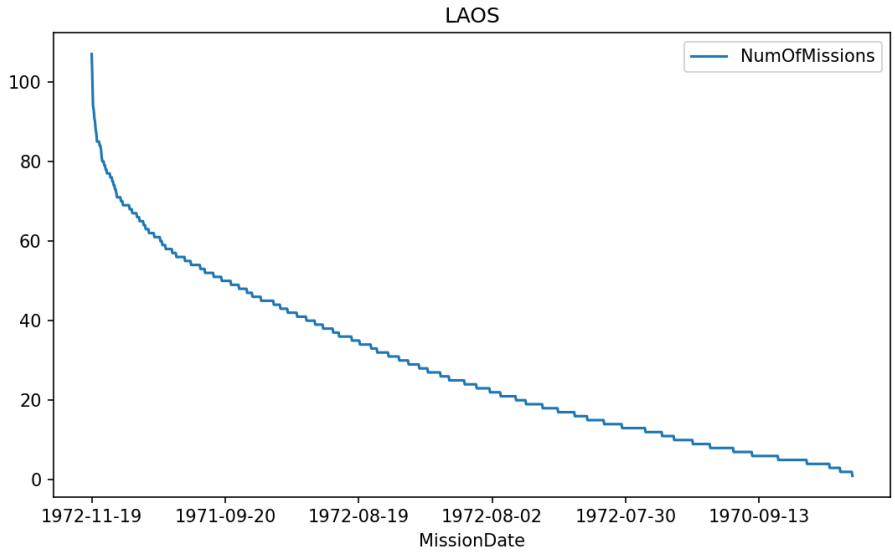
```
>>> Bombing_Operations.createOrReplaceTempView("Bombing_Operations")
>>> query = """ select ContryFlyingMission, MissionDate, count(*) as NumOfMissions from Bombing_Operations group by ContryFlyi
ngMission,MissionDate order by NumOfMissions desc"""
>>> missionCount1=spark.sql(query)
>>> missionCount1.show()
+--------------------+-----------+-------------+
```

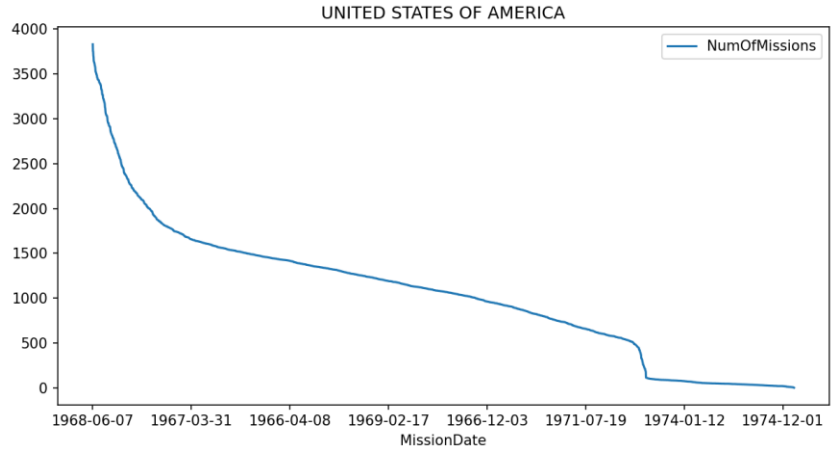
```
+--------------------+-----------+-------------+
| ContryFlyingMission|MissionDate|NumOfMissions|
+--------------------+-----------+-------------+
|UNITED STATES OF ...| 1968-06-07|         3832|
|UNITED STATES OF ...| 1968-09-16|         3755|
|UNITED STATES OF ...| 1968-09-18|         3752|
|UNITED STATES OF ...| 1968-09-17|         3706|
|UNITED STATES OF ...| 1968-05-11|         3692|
|UNITED STATES OF ...| 1968-09-14|         3646|
|UNITED STATES OF ...| 1968-06-16|         3639|
|UNITED STATES OF ...| 1968-06-23|         3630|
|UNITED STATES OF ...| 1968-06-11|         3628|
|UNITED STATES OF ...| 1968-09-19|         3614|
|UNITED STATES OF ...| 1968-05-25|         3605|
|UNITED STATES OF ...| 1968-05-22|         3590|
|UNITED STATES OF ...| 1968-06-19|         3582|
|UNITED STATES OF ...| 1968-09-15|         3568|
|UNITED STATES OF ...| 1968-05-13|         3531|
|UNITED STATES OF ...| 1968-05-26|         3530|
|UNITED STATES OF ...| 1968-09-28|         3517|
|UNITED STATES OF ...| 1968-09-08|         3511|
|UNITED STATES OF ...| 1968-05-19|         3502|
|UNITED STATES OF ...| 1968-06-24|         3501|
+--------------------+-----------+-------------+
only showing top 20 rows
```

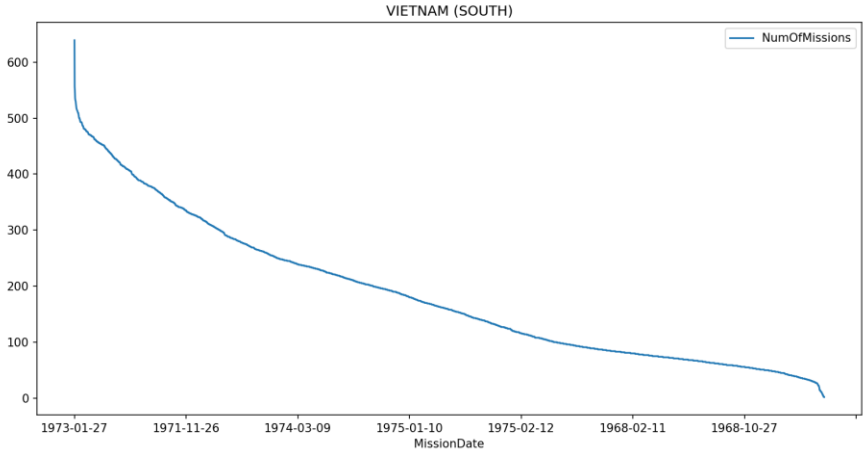
```
>>>
>>> missionCount1=missionCount1.toPandas()
>>>
>>> grouped = missionCount1.groupby('ContryFlyingMission')
>>>
>>> for name,group in grouped:
...     group.plot( x='MissionDate', y='NumOfMissions',title=name)
...
<AxesSubplot:title={'center':'AUSTRALIA'}, xlabel='MissionDate'>
<AxesSubplot:title={'center':'KOREA (SOUTH)'}, xlabel='MissionDate'>
<AxesSubplot:title={'center':'LAOS'}, xlabel='MissionDate'>
<AxesSubplot:title={'center':'UNITED STATES OF AMERICA'}, xlabel='MissionDate'>
<AxesSubplot:title={'center':'VIETNAM (SOUTH)'}, xlabel='MissionDate'>
>>> plt.show()
```




LAOS
NumOfMissions
100
80
60
40
20
0
1972-11-19
1971-09-20
1972-08-19
1972-08-02
1972-07-30
1970-09-13
MissionDate


UNITED STATES OF AMERICA
NumOfMissions
4000
3500
3000
2500
2000
1500
1000
500
0
1968-06-07
1967-03-31
1966-04-08
1969-02-17
1966-12-03
1971-07-19
1974-01-12
1974-12-01
MissionDate


VIETNAM (SOUTH)
NumOfMissions
600
500
400
300
200
100
0
1973-01-27
1971-11-26
1974-03-09
1975-01-10
1975-02-12
1968-02-11
1968-10-27
MissionDate

## Problem 2:

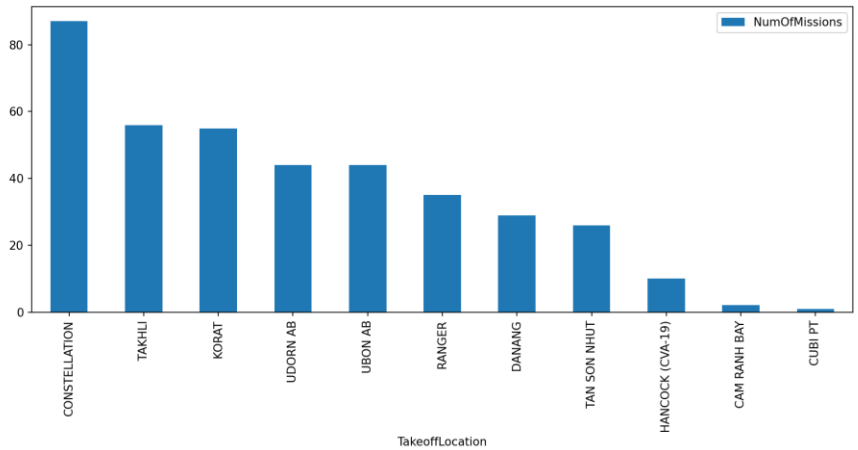
## Find the number of missions completed for each take-off location on June 29, 1966 by the US Air Force over NORTH VIETNAM.

## Codes and Outputs:

```
>>> Bombing_Operations.createOrReplaceTempView("Bombing_Operations")
>>> query = """ select TakeoffLocation, count(*) as NumOfMissions from Bombing_Operations where MissionDate='1966-06-29' and C
ontryFlyingMission='UNITED STATES OF AMERICA' and TargetCountry= 'NORTH VIETNAM' group by TakeoffLocation order by NumOfMissio
ns desc"""
>>> missionCount1=spark.sql(query)
>>> missionCount1.show()
```

```
+----------------+-------------+
| TakeoffLocation|NumOfMissions|
+----------------+-------------+
|   CONSTELLATION|           87|
|          TAKHLI|           56|
|           KORAT|           55|
|        UDORN AB|           44|
|         UBON AB|           44|
|          RANGER|           35|
|          DANANG|           29|
|    TAN SON NHUT|           26|
|HANCOCK (CVA-19)|           10|
|    CAM RANH BAY|            2|
|         CUBI PT|            1|
+----------------+-------------+
```

```
>>> missions_count1_pd= missionCount1.toPandas()
>>> missions_count1_pd.plot(kind="bar", x="TakeoffLocation", y="NumOfMissions")
<AxesSubplot:xlabel='TakeoffLocation'>
>>> plt.show()
```



## Problem 3:

**Find the number of times each AirCraftType used during the Vietnam war. (Need join operation)**

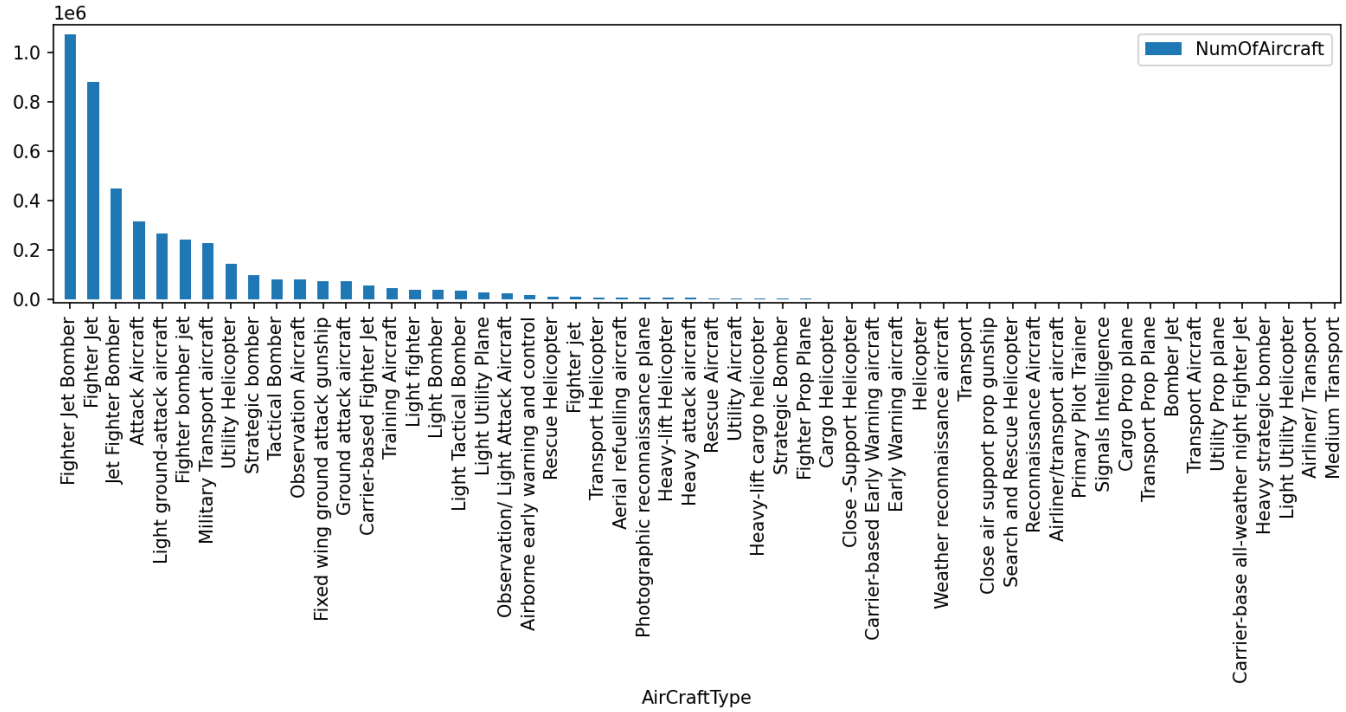
## Codes and Outputs:

```
>>> query = Bombing_Operations.join(Aircraft_Glossary,'AirCraft')
>>> query.createOrReplaceTempView("query")
>>> query2= """ select AirCraftType, count(*) as NumOfAircraft from query group by AirCraftType order by NumOfAircraft desc"""

>>> missionCount1=spark.sql(query2)
>>> missionCount1.show()
```

```
+--------------------+-------------+
|        AirCraftType|NumOfAircraft|
+--------------------+-------------+
|  Fighter Jet Bomber|      1073126|
|         Fighter Jet|       882594|
|  Jet Fighter Bomber|       451385|
|     Attack Aircraft|       315246|
|Light ground-atta...|       267457|
|  Fighter bomber jet|       242231|
|Military Transpor...|       228426|
|  Utility Helicopter|       146653|
|    Strategic bomber|        99100|
|     Tactical Bomber|        82219|
|Observation Aircraft|        81820|
|Fixed wing ground...|        75058|
|Ground attack air...|        73843|
|Carrier-based Fig...|        58691|
|   Training Aircraft|        48435|
|       Light fighter|        39999|
|        Light Bomber|        39262|
|Light Tactical Bo...|        34738|
| Light Utility Plane|        28582|
|Observation/ Ligh...|        24491|
+--------------------+-------------+
only showing top 20 rows
```

```
>>> missions_count1_pd= missionCount1.toPandas()
>>> missions_count1_pd.plot(kind="bar", x="AirCraftType", y="NumOfAircraft")
<AxesSubplot:xlabel='AirCraftType'>
>>> plt.show()
```


1e6
1.0
0.8
0.6
0.4
0.2
0.0
NumOfAircraft
Fighter Jet Bomber
Fighter Jet
Jet Fighter Bomber
Attack Aircraft
Light ground-attack aircraft
Fighter bomber jet
Military Transport aircraft
Utility Helicopter
Strategic bomber
Tactical Bomber
Observation Aircraft
Fixed wing ground attack gunship
Ground attack aircraft
Carrier-based Fighter Jet
Training Aircraft
Light fighter
Light Bomber
Light Tactical Bomber
Light Utility Plane
Observation/ Light Attack Aircraft
Airborne early warning and control
Rescue Helicopter
Fighter jet
Transport Helicopter
Aerial refuelling aircraft
Photographic reconnaissance plane
Heavy-lift Helicopter
Heavy attack aircraft
Rescue Aircraft
Utility Aircraft
Heavy-lift cargo helicopter
Strategic Bomber
Fighter Prop Plane
Cargo Helicopter
Close -Support Helicopter
Carrier-based Early Warning aircraft
Early Warning aircraft
Helicopter
Weather reconnaissance aircraft
Transport
Close air support prop gunship
Search and Rescue Helicopter
Reconnaissance Aircraft
Airliner/transport aircraft
Primary Pilot Trainer
Signals Intelligence
Cargo Prop plane
Transport Prop Plane
Bomber Jet
Transport Aircraft
Utility Prop plane
Carrier-base all-weather night Fighter Jet
Heavy strategic bomber
Light Utility Helicopter
Airliner/ Transport
Medium Transport
AirCraftType