In [1]:

```
import findspark
findspark.init()
findspark.find()
```

Out[1]:

```
'C:\\Spark\\spark-3.0.3-bin-hadoop2.7'
```

In [2]:

```

from pyspark.sql import SparkSession
spark = SparkSession.builder.getOrCreate()
df = spark.sql("select 'spark' as hello ")
df.show()
# spark.stop()
```

```
+-----+
|hello|
+-----+
|spark|
+-----+
```

Admin_Data_Uber

In [72]:

```
from pyspark.sql import *
p = spark.read.csv(r"E:\Jupyter Noteii\Python Basics\Revature\P2\DS\Main_DS\Admi
p.show(0)
```

```
+----------+--------+----+------+---+---------+------+-----------+-----------+
-----+-----------+-------------+-------+----+------+--------+--------+----+---
----------+---------+---------+---------+----------+----------+-----------+---
-----------+---------------+---------+----------+------------+--------+-------
---+------+
|Start_time|End_time|Name|Mobile|Age|Pin_Codes|Source|Vaccine_cus|Destination|
Miles|Est_Costing|Ride_category|Purpose|temp|clouds|pressure|humidity|wind|acc
quire_vehi|free_vehi|Lattitute|Longitude|locationID|rating_cus|Riders_Name|Rid
ers_contact|Trusted_Contact|Rating_RI|Vaccine_Ri|Payment_mode|Discount|Final_c
ost|Status|
+----------+--------+----+------+---+---------+------+-----------+-----------+
-----+-----------+-------------+-------+----+------+--------+--------+----+---
----------+---------+---------+---------+----------+----------+-----------+---
-----------+---------------+---------+----------+------------+--------+-------
---+------+
+----------+--------+----+------+---+---------+------+-----------+-----------+
-----+-----------+-------------+-------+----+------+--------+--------+----+---
----------+---------+---------+---------+----------+----------+-----------+---
-----------+---------------+---------+----------+------------+--------+-------
---+------+
only showing top 0 rows
```

Using TempView and Spark.sql

In [ ]:

In [73]:

```
p.createOrReplaceTempView("pa")
```

In [74]:
```
spark.sql("describe pa").show()
```

```
+-------------+---------+-------+
|     col_name|data_type|comment|
+-------------+---------+-------+
|   Start_time|   string|   null|
|     End_time|   string|   null|
|         Name|   string|   null|
|       Mobile|   string|   null|
|          Age|   string|   null|
|    Pin_Codes|   string|   null|
|       Source|   string|   null|
|  Vaccine_cus|   string|   null|
|  Destination|   string|   null|
|        Miles|   string|   null|
|  Est_Costing|   string|   null|
|Ride_category|   string|   null|
|      Purpose|   string|   null|
|         temp|   string|   null|
|       clouds|   string|   null|
|     pressure|   string|   null|
|     humidity|   string|   null|
|         wind|   string|   null|
|accquire_vehi|   string|   null|
|    free_vehi|   string|   null|
+-------------+---------+-------+
only showing top 20 rows
```

In [76]:
```
spark.sql("select count(Pin_Codes) from pa").show(5)
```

```
+----------------+
|count(Pin_Codes)|
+----------------+
|             500|
+----------------+
```

In [88]:
```
# spark.sql("select * from pa").show(2)
```

In [91]:
```
spark.sql("select Source,Destination,Miles from pa where Miles=(select max(Miles
```

```
+----------------+-----------+-----+
|          Source|Destination|Miles|
+----------------+-----------+-----+
|     East Harlem|       Cary|  9.9|
|         Midtown|     Durham|  9.9|
|     East Harlem|       Cary|  9.9|
|Flatiron District|     Durham|  9.9|
|         Midtown|       Cary|  9.9|
+----------------+-----------+-----+
only showing top 5 rows
```

In [92]:
```
spark.sql("select Source,Destination,Miles from pa where Miles=(select min(Miles
```

```
+-----------------+-----------+-----+
|           Source|Destination|Miles|
+-----------------+-----------+-----+
|  West Palm Beach|       Cary|  0.5|
|Flatiron District| Katunayaka|  0.5|
|      Fort Pierce|    Tribeca|  0.5|
+-----------------+-----------+-----+
```

In [97]:
```
spark.sql("select Source,Ride_category,count(Ride_category) as priority from pa
```

```
+------+-------------+--------+
|Source|Ride_category|priority|
+------+-------------+--------+
|  Cary|        Prime|      12|
|  Cary|         Auto|      12|
|  Cary|    Uber-Mini|      12|
|  Cary|   Uber-Micro|       8|
|  Cary|         Bike|       8|
+------+-------------+--------+
only showing top 5 rows
```

In [4]:

```
p.printSchema()
```

```
root
 |-- Start_time: string (nullable = true)
 |-- End_time: string (nullable = true)
 |-- Name: string (nullable = true)
 |-- Mobile: string (nullable = true)
 |-- Age: string (nullable = true)
 |-- Pin-Codes: string (nullable = true)
 |-- Source: string (nullable = true)
 |-- Vaccine_cus: string (nullable = true)
 |-- Destination: string (nullable = true)
 |-- Miles: string (nullable = true)
 |-- Est_Costing: string (nullable = true)
 |-- Ride_category: string (nullable = true)
 |-- Purpose: string (nullable = true)
 |-- temp: string (nullable = true)
 |-- clouds: string (nullable = true)
 |-- pressure: string (nullable = true)
 |-- humidity: string (nullable = true)
 |-- wind: string (nullable = true)
 |-- accquire_vehi: string (nullable = true)
 |-- free_vehi: string (nullable = true)
 |-- Lattitute: string (nullable = true)
 |-- Longitude: string (nullable = true)
 |-- locationID: string (nullable = true)
 |-- rating_cus: string (nullable = true)
 |-- Riders_Name: string (nullable = true)
 |-- Riders_contact: string (nullable = true)
 |-- Trusted_Contact: string (nullable = true)
 |-- Rating_RI: string (nullable = true)
 |-- Vaccine_Ri: string (nullable = true)
 |-- Payment_mode: string (nullable = true)
 |-- Discount: string (nullable = true)
 |-- Final_cost: string (nullable = true)
 |-- Status: string (nullable = true)
```

In [8]:

```
#1- print the start_time, end_time, name, mobile,age,free_vehicle status ?
p.select("Start_time","End_time","Name","Mobile","Age","free_vehi").show(2)
```

```
+--------------+--------------+-------+----------+---+---------+
|    Start_time|      End_time|   Name|    Mobile|Age|free_vehi|
+--------------+--------------+-------+----------+---+---------+
|1/1/2016 21:11|1/1/2016 21:17| Almire|9298608912| 21|       17|
|1/2/2016 20:25|1/2/2016 20:38|Frazier|8621617385| 27|       24|
+--------------+--------------+-------+----------+---+---------+
only showing top 2 rows
```

In [14]:
```
# 2 - print many columns where Purpose of the Ride is "Meeting" & Rating_RI is 5
p.select("Start_time","End_time","Name","Mobile","Age").filter((p.Purpose=='Meet
```

```
+---------------+---------------+------+----------+---+
|     Start_time|       End_time|  Name|    Mobile|Age|
+---------------+---------------+------+----------+---+
|1/29/2016 13:24|1/29/2016 13:47|Aubert|9524013920| 58|
|2/13/2016 14:21|2/13/2016 14:41|   Val|8472948148| 20|
+---------------+---------------+------+----------+---+
only showing top 2 rows
```

In [21]:
```
# 3 - print many columns where purpose = Meeting, rider rating > 3 and vaccine s
p.select("Start_time","End_time","Name","Mobile","Age").filter((p.Purpose=='Meet
```

```
+---------------+---------------+-------+----------+---+
|     Start_time|       End_time|   Name|    Mobile|Age|
+---------------+---------------+-------+----------+---+
| 1/5/2016 17:31| 1/5/2016 17:45| Editha|9954004976| 20|
|1/10/2016 19:12|1/10/2016 19:32|Carlyle|8333928562| 22|
+---------------+---------------+-------+----------+---+
only showing top 2 rows
```

In [18]:
```
tails where purpose = Meeting, Rating > 3, Discount = 10%
ime","End_time","Name","Mobile","Age","Payment_mode").filter((p.Purpose=='Meetin
```

```
+---------------+---------------+--------+----------+---+------------+
|     Start_time|       End_time|    Name|    Mobile|Age|Payment_mode|
+---------------+---------------+--------+----------+---+------------+
|1/28/2016 15:11|1/28/2016 15:31|    Berk|8875035370| 30| Uber wallet|
|1/29/2016 10:56|1/29/2016 11:07|Valentia|9958471700| 42| Uber wallet|
+---------------+---------------+--------+----------+---+------------+
only showing top 2 rows
```

In [83]:
```
# p.printSchema()
```

In [78]:
```
","clouds","humidity","wind","Lattitute","Longitude","locationID").filter((p.Sta
```

```
+--------+-----+--------+------+--------+-----+---------+---------+----------+
|    Name| temp|pressure|clouds|humidity| wind|Lattitute|Longitude|locationID|
+--------+-----+--------+------+--------+-----+---------+---------+----------+
|Zedekiah|43.27|   990.8|   0.8|    0.71|  8.3|  40.7271| -73.9803|       229|
| Maurice| 43.2|  990.79|   0.8|    0.71| 8.31|   40.758| -73.9761|       188|
|   Janey|41.95|  991.63|  0.81|    0.73|10.87|  40.7531| -74.0039|       224|
| Huntlee|43.05|  990.82|  0.81|    0.72| 8.31|  40.7389| -74.0393|       238|
| Pauline|39.23|  996.09|  0.83|    0.66|10.67|  40.9859| -74.1578|       230|
+--------+-----+--------+------+--------+-----+---------+---------+----------+
only showing top 5 rows
```

In [79]:
```
',"Lattitute","Longitude","locationID").filter((p.Status == "Assigned")&(p.clouds
```

```
+-------+-----+--------+------+--------+-----+---------+---------+----------+
|   Name| temp|pressure|clouds|humidity| wind|Lattitute|Longitude|locationID|
+-------+-----+--------+------+--------+-----+---------+---------+----------+
|Maurice| 43.2|  990.79|   0.8|    0.71| 8.31|   40.758| -73.9761|       188|
|  Janey|41.95|  991.63|  0.81|    0.73|10.87|  40.7531| -74.0039|       224|
|Huntlee|43.05|  990.82|  0.81|    0.72| 8.31|  40.7389| -74.0393|       238|
|Pauline|39.23|  996.09|  0.83|    0.66|10.67|  40.9859| -74.1578|       230|
|Geordie|39.56|  996.07|  0.83|    0.66|10.79|  40.7336|   -73.99|        87|
+-------+-----+--------+------+--------+-----+---------+---------+----------+
only showing top 5 rows
```

In [80]:
```
cationID").filter((p.Status == "Assigned")&(p.clouds < 1)&(p.Discount == "10%")&
```

```
+---------+-----+--------+------+--------+-----+---------+---------+----------
+
|     Name| temp|pressure|clouds|humidity| wind|Lattitute|Longitude|locationID
|
+---------+-----+--------+------+--------+-----+---------+---------+----------
+
|  Maurice| 43.2|  990.79|   0.8|    0.71| 8.31|   40.758| -73.9761|       188
|
|  Pauline|39.23|  996.09|  0.83|    0.66|10.67|  40.9859| -74.1578|       230
|
|    Gusta|27.34|  1033.4|  0.15|    0.81| 3.04|  40.7437| -73.9985|       125
|
|    Alena|45.98| 1021.65|  0.99|     0.9| 5.64|  40.7249| -74.0355|       147
|
|Annadiane|   37| 1001.76|  0.29|    0.67| 10.1|   40.771|  -73.866|        79
|
+---------+-----+--------+------+--------+-----+---------+---------+----------
+
only showing top 5 rows
```

In [81]:
```
ted_Contact","Rating_RI","Vaccine_Ri","Final_cost").filter((p.Payment_mode == "Gp
```

```
+-----------+--------------+---------------+---------+----------+----------+
|Riders_Name|Riders_contact|Trusted_Contact|Rating_RI|Vaccine_Ri|Final_cost|
+-----------+--------------+---------------+---------+----------+----------+
|    Johanna|    9181026109|            YES|      3.6|       YES|     49.98|
|      Amara|    9247349792|            YES|      3.6|       YES|       381|
|      Price|    9647090347|            YES|      4.5|       YES|     79.38|
|       Burk|    9891913387|             NO|        5|        NO|        39|
|      Price|    9647090347|            YES|      4.5|       YES|      1710|
+-----------+--------------+---------------+---------+----------+----------+
only showing top 5 rows
```

In [82]:
```
usted_Contact","Rating_RI","Vaccine_Ri","Final_cost","Status").filter((p.Payment_
```

```
+-----------+-------------+--------------+---------+----------+----------+--
------+
|Riders_Name|Riders_contact|Trusted_Contact|Rating_RI|Vaccine_Ri|Final_cost|
Status|
+-----------+-------------+--------------+---------+----------+----------+--
------+
|  Siegfried|    9419083292|           YES|      4.8|       YES|     256.5| A
rrived|
|      Cindi|    9080785128|            NO|        5|        NO|     280.5| A
rrived|
|    Ellette|    8949025277|            NO|      4.8|        NO|       705|As
signed|
|       Mair|    9111821429|           YES|      3.6|       YES|       693|As
signed|
|    Teriann|    9115587792|           YES|      4.5|       YES|       495|As
signed|
+-----------+-------------+--------------+---------+----------+----------+--
------+
only showing top 5 rows
```

In [87]:
```
_Ri","Final_cost","Status").filter((p.Payment_mode == "Gpay")&(p.Miles > 15)).or
```

```
+-----------+-------------+--------------+---------+----------+----------+--
------+
|Riders_Name|Riders_contact|Trusted_Contact|Rating_RI|Vaccine_Ri|Final_cost|
Status|
+-----------+-------------+--------------+---------+----------+----------+--
------+
|      Cindi|    9080785128|            NO|        5|        NO|     280.5| A
rrived|
|  Siegfried|    9419083292|           YES|      4.8|       YES|     256.5| A
rrived|
|     Haskel|    9199720906|            NO|      4.8|        NO|       340|As
signed|
|    Ellette|    8949025277|            NO|      4.8|        NO|       705|As
signed|
|    Teriann|    9115587792|           YES|      4.5|       YES|       495|As
signed|
+-----------+-------------+--------------+---------+----------+----------+--
------+
only showing top 5 rows
```

In [102]:
```
filter((p.Payment_mode == "Gpay")&(p.Miles > 15)).orderBy((p.Rating_RI.desc()),(
```

```
+-----------+--------------+---------------+---------+----------+----------+--
------+
|Riders_Name|Riders_contact|Trusted_Contact|Rating_RI|Vaccine_Ri|Final_cost|
Status|
+-----------+--------------+---------------+---------+----------+----------+--
------+
|      Cindi|    9080785128|             NO|        5|        NO|     280.5| A
rrived|
|    Ellette|    8949025277|             NO|      4.8|        NO|       705|As
signed|
|     Haskel|    9199720906|             NO|      4.8|        NO|       340|As
signed|
|  Siegfried|    9419083292|            YES|      4.8|       YES|     256.5| A
rrived|
|       Moss|    9651245558|            YES|      4.5|       YES|       690|As
signed|
+-----------+--------------+---------------+---------+----------+----------+--
------+
only showing top 5 rows
```

In [ ]:
```

```

In [ ]:
```

```

Custumer_data

In [23]:

```
a=spark.read.csv('E:\Jupyter Noteii\Python Basics\Revature\P2\DS\Customer_table.
a.show(2)
```

```
+--------------+--------------+-------+----------+---+---------+-----------+--
---------+-----+-----------+-------------+---------------+-----------+--------
--+---------------+---------+----------+------------+--------+----------+-----
----+
|    Start_time|      End_time|   Name|    Mobile|Age|Pin-Codes|     Source|De
stination|Miles|Est_Costing|Ride_category|        Purpose|Riders_Name|  Rid_co
nt|Trusted_Contact|Rating_RI|Vaccine_Ri|Payment_mode|Discount|Final_cost|  Rid
e_St|
+--------------+--------------+-------+----------+---+---------+-----------+--
---------+-----+-----------+-------------+---------------+-----------+--------
--+---------------+---------+----------+------------+--------+----------+-----
----+
|1/1/2016 21:11|1/1/2016 21:17| Almire|9298608912| 21|   318886|Fort Pierce|Fo
rt Pierce|  5.1|      49.98|         Bike| Meal/Entertain|    Johanna|91810261
09|            YES|      3.6|       YES|        Gpay|      0%|     49.98|Cance
lled|
|1/2/2016 20:25|1/2/2016 20:38|Frazier|8621617385| 27|   578822|Fort Pierce|Fo
rt Pierce|  4.8|      47.04|         Bike|Errand/Supplies|    Charlot|98554031
24|            YES|      4.5|       YES|      Phonpe|      0%|     47.04|   Arr
ived|
+--------------+--------------+-------+----------+---+---------+-----------+--
---------+-----+-----------+-------------+---------------+-----------+--------
--+---------------+---------+----------+------------+--------+----------+-----
----+
only showing top 2 rows
```

In [24]:

```
a.printSchema()
```

```
root
 |-- Start_time: string (nullable = true)
 |-- End_time: string (nullable = true)
 |-- Name: string (nullable = true)
 |-- Mobile: long (nullable = true)
 |-- Age: integer (nullable = true)
 |-- Pin-Codes: integer (nullable = true)
 |-- Source: string (nullable = true)
 |-- Destination: string (nullable = true)
 |-- Miles: double (nullable = true)
 |-- Est_Costing: double (nullable = true)
 |-- Ride_category: string (nullable = true)
 |-- Purpose: string (nullable = true)
 |-- Riders_Name: string (nullable = true)
 |-- Rid_cont: long (nullable = true)
 |-- Trusted_Contact: string (nullable = true)
 |-- Rating_RI: double (nullable = true)
 |-- Vaccine_Ri: string (nullable = true)
 |-- Payment_mode: string (nullable = true)
 |-- Discount: string (nullable = true)
 |-- Final_cost: double (nullable = true)
 |-- Ride_St: string (nullable = true)
```

```
In [26]: a.select("Ride_category","Purpose","Payment_mode").filter(a.Ride_St == 'Assigned
```

```
+-------------+--------------+------------+
|Ride_category|       Purpose|Payment_mode|
+-------------+--------------+------------+
|         Bike|       Meeting|       Paytm|
|         Bike|Customer Visit| Uber wallet|
|         Bike|Meal/Entertain|        cash|
|         Bike|       Meeting| Uber wallet|
|         Bike|       Meeting| Uber wallet|
+-------------+--------------+------------+
only showing top 5 rows
```

Rider data set

```
In [100]: b=spark.read.csv('E:\Jupyter Noteii\Python Basics\Revature\P2\DS\Riders_data.csv
          b.show(2)
```

```
+--------------+--------------+-------+----------+---+---------+-----------+--
---------+-----+-----------+-------------+-----------+----------+----------+--
---------+---------------+-----------+------------+--------+----------+-------
--+
|    Start_time|      End_time|   Name|    Mobile|Age|Pin-Codes|     Source|De
stination|Miles|Est_Costing|Ride_category|Riders_Name|  Rid_cont|Rating_cus|Lo
cation_Id|Trusted_Contact|Vaccine_Cus|Payment_mode|Discount|Final_cost|  Ride_
St|
+--------------+--------------+-------+----------+---+---------+-----------+--
---------+-----+-----------+-------------+-----------+----------+----------+--
---------+---------------+-----------+------------+--------+----------+-------
--+
|1/1/2016 21:11|1/1/2016 21:17| Almire|9298608912| 21|   318886|Fort Pierce|Fo
rt Pierce|  5.1|      49.98|         Bike|    Johanna|9181026109|         3|
141|            YES|        YES|        Gpay|      0%|     49.98|Cancelled|
|1/2/2016 20:25|1/2/2016 20:38|Frazier|8621617385| 27|   578822|Fort Pierce|Fo
rt Pierce|  4.8|      47.04|         Bike|    Charlot|9855403124|         4|
65|            YES|        YES|      Phonpe|      0%|     47.04|  Arrived|
+--------------+--------------+-------+----------+---+---------+-----------+--
---------+-----+-----------+-------------+-----------+----------+----------+--
---------+---------------+-----------+------------+--------+----------+-------
--+
only showing top 2 rows
```

In [101]:
```
b.printSchema()
```

```
root
 |-- Start_time: string (nullable = true)
 |-- End_time: string (nullable = true)
 |-- Name: string (nullable = true)
 |-- Mobile: long (nullable = true)
 |-- Age: integer (nullable = true)
 |-- Pin-Codes: integer (nullable = true)
 |-- Source: string (nullable = true)
 |-- Destination: string (nullable = true)
 |-- Miles: double (nullable = true)
 |-- Est_Costing: double (nullable = true)
 |-- Ride_category: string (nullable = true)
 |-- Riders_Name: string (nullable = true)
 |-- Rid_cont: long (nullable = true)
 |-- Rating_cus: integer (nullable = true)
 |-- Location_Id: integer (nullable = true)
 |-- Trusted_Contact: string (nullable = true)
 |-- Vaccine_Cus: string (nullable = true)
 |-- Payment_mode: string (nullable = true)
 |-- Discount: string (nullable = true)
 |-- Final_cost: double (nullable = true)
 |-- Ride_St: string (nullable = true)
```

In [107]:
```
b.createOrReplaceTempView("rider")
```

In [111]:
```
spark.sql("select distinct(source) from rider").show(5)
```

```
+--------------------+
|              source|
+--------------------+
|Pontchartrain Shores|
|            Fairmont|
|        Briar Meadow|
|          Menlo Park|
|           Palo Alto|
+--------------------+
only showing top 5 rows
```

In [110]:
```
spark.sql("select Location_Id,Est_Costing from rider where Est_Costing<100").sho
```

```
+-----------+-----------+
|Location_Id|Est_Costing|
+-----------+-----------+
|        141|      49.98|
|         65|      47.04|
|        100|      46.06|
|         90|      42.14|
|        228|      69.58|
+-----------+-----------+
only showing top 5 rows
```

In [ ]:
```
```