

```
!pip install graphframes
from graphframes import *
```

```
Requirement already satisfied: graphframes in /databricks/python3/lib/python3.8/site-packages (0.6)
Requirement already satisfied: numpy in /databricks/python3/lib/python3.8/site-packages (from graphframes) (1.19.2)
Requirement already satisfied: nose in /databricks/python3/lib/python3.8/site-packages (from graphframes) (1.3.7)
WARNING: You are using pip version 21.0.1; however, version 22.0.4 is available.
You should consider upgrading via the '/databricks/python3/bin/python -m pip install --upgrade pip' command.
```

```
stations = sc.textFile("FileStore/tables/week3/stations.csv")
stations = stations.map(lambda x: x.split(","))
stations = stations.map(lambda x: (int(x[0]), str(x[1])))
vertices = sqlContext.createDataFrame(stations, ["id", "station_name"])
vertices.show(3)
```

```
+---+-----+
| id|      station_name|
+---+-----+
|  2|San Jose Diridon ...|
|  3|San Jose Civic Ce...|
|  4|Santa Clara at Al...|
+---+-----+
only showing top 3 rows
```

```
trips = sc.textFile("FileStore/tables/week3/trips.csv")
trips = trips.map(lambda x: x.split(","))
trips = trips.map(lambda x: (int(x[4]), int(x[7])))
edges = sqlContext.createDataFrame(trips, ["src", "dst"])
edges.show(3)
```

```
+---+---+
|src|dst|
+---+---+
| 50| 70|
| 31| 27|
| 47| 64|
+---+---+
```

only showing top 3 rows

```
graph = GraphFrame(vertices, edges)
```

```
graph.inDegrees.sort("inDegree", ascending=False).show(3)
```

```
+---+-----+
| id|inDegree|
+---+-----+
| 70|   34810|
| 69|   22523|
| 50|   17810|
+---+-----+
```

only showing top 3 rows

```
graph.outDegrees.sort("outDegree", ascending=False).show(3)
```

```
+---+-----+
| id|outDegree|
+---+-----+
| 70|    26304|
| 69|    21758|
| 50|    17255|
+---+-----+
```

only showing top 3 rows

```
+---+-----+-----+
| id|      station_name|      pagerank|
+---+-----+-----+
|  2|San Jose Diridon ...|  3.95311114650878|
| 70|San Francisco Cal...|  3.3568798329621|
| 28|Mountain View Cal...|2.4533269567415403|
+---+-----+-----+
```

only showing top 3 rows

