Taller Spark – DataFramas

1)

Descargar el dataset de taxis de NYC para el mes de junio:

<https://nyc-tlc.s3.amazonaws.com/trip+data/yellow_tripdata_2019-06.csv>

Descargar archivo de zonas:

<https://nyc-tlc.s3.amazonaws.com/misc/taxi+_zo> [ne\_lookup.csv](https://nyc-tlc.s3.amazonaws.com/misc/taxi+_zone_lookup.csv)

2) Escribir los respectivos programas en Spark usando Dataframes para realizar las siguientes consultas:

1. Distancia Promedio de un recorrigo en Taxi NYC
2. Cuántas formas de pago hay?
3. Taxi con mayor número de viajes
4. Listar número de viajes por dia en el mes de junio de 2019
5. Área donde mayor número de pasajeros se recoge
6. Número de viajes que se dirigieron al “Bronx”
7. Número promedio de personas por viaje que se dirigen al aeropuerto JFK
8. Distancia y Costo promedio de tomar un taxi del Aeropuerto JFK a Manhattan Valley
9. Cual es el recorrido más frencuente(entre qué zona y qué zona)
10. Tiempo Promedio de un viaje.