

### Visualization of question c:-

```
g=appended3.groupBy(weekofyear("tpep_pickup_datetime").alias("date_by_week"),"VendorID").agg  
g(sum("distance")).orderBy("date_by_week")
```

```
display(g)
```



### Visualization of Question d:-

```
es=appended3.select("VendorID","distance");
```

```
es.show(100)
```

```
from pyspark.sql.functions import sum as _sum
```

```
es=es.groupBy('VendorID').agg(_sum('distance').alias('Sum_distance'))
```

```
display(es)
```

