**Practice – EDA Visualisation**

**Visualisations in Databricks**

**%sql**

USE CATALOG samples;

SELECT

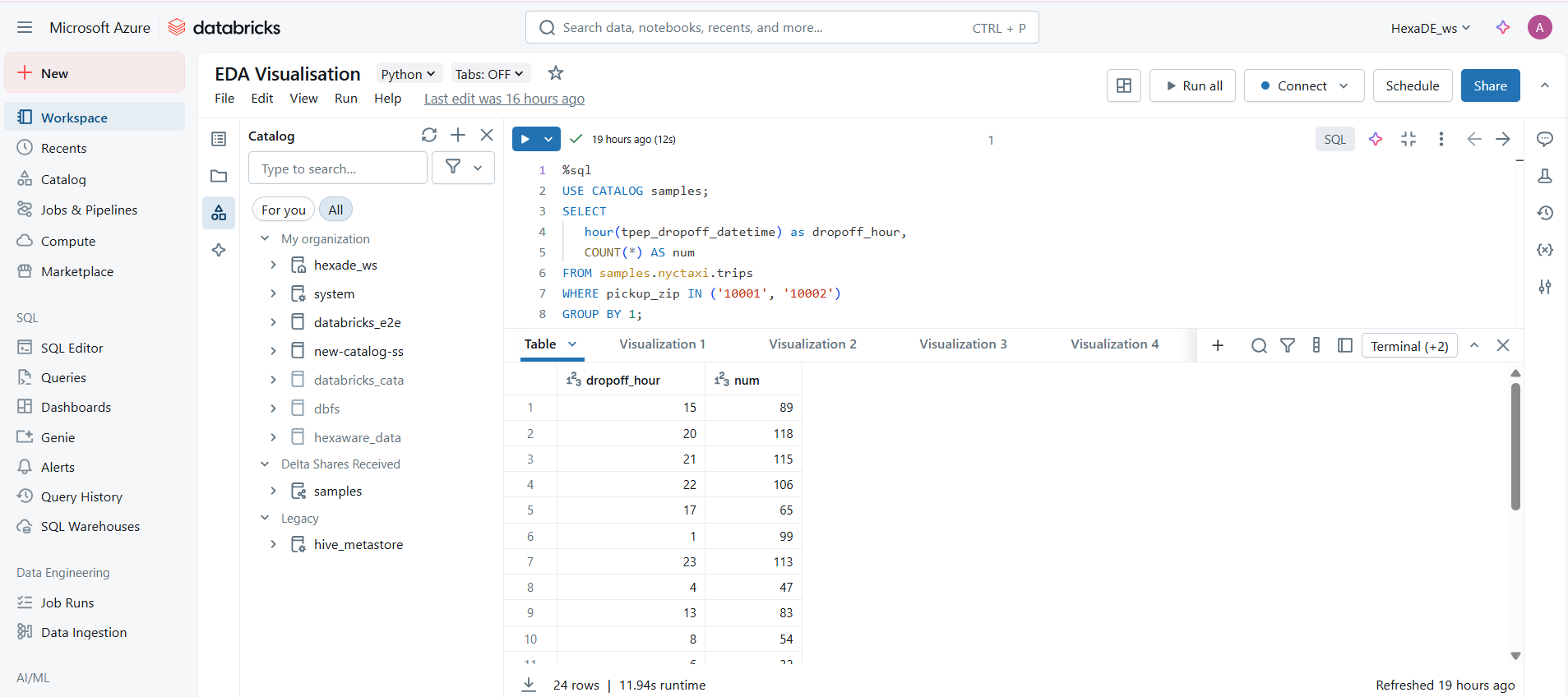
   hour(tpep\_dropoff\_datetime) as dropoff\_hour,

   COUNT(\*) AS num

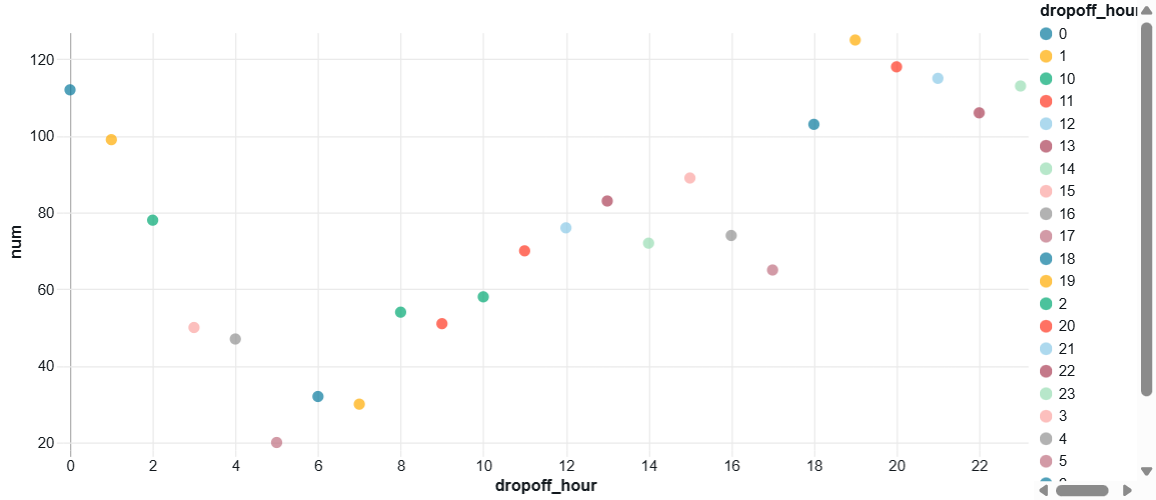
FROM samples.nyctaxi.trips

WHERE pickup\_zip IN ('10001', '10002')

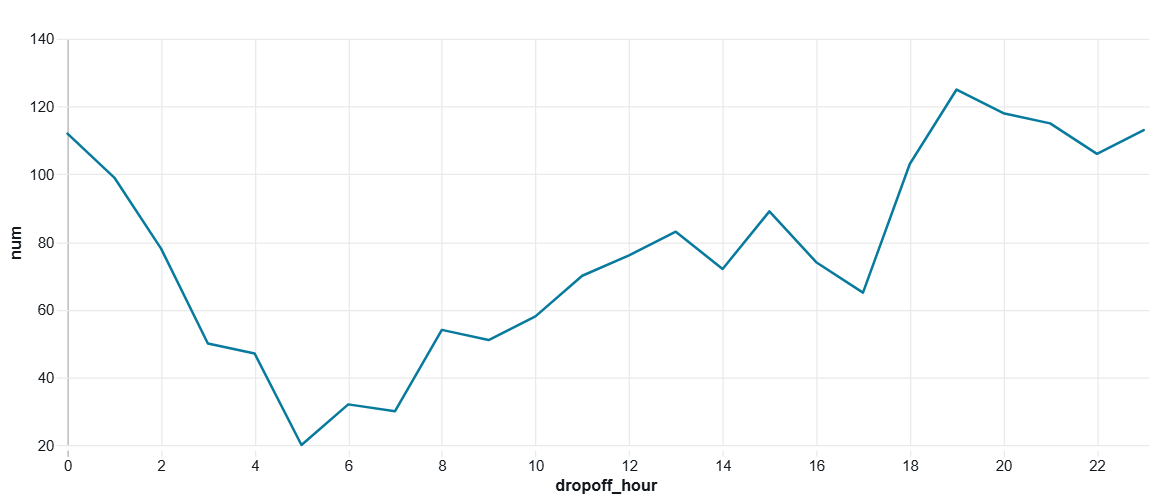
GROUP BY 1;

****

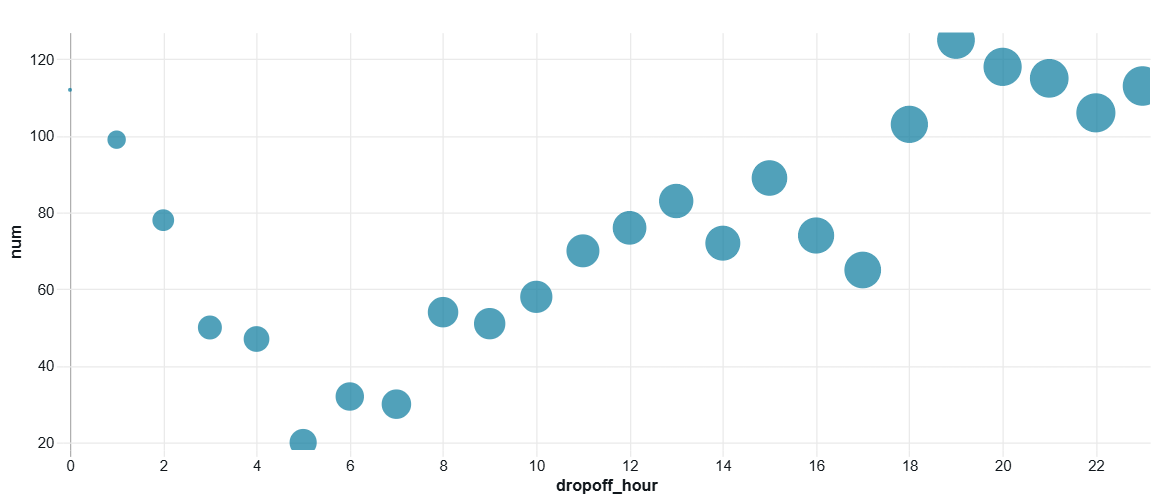
**Visualisations Types**

****

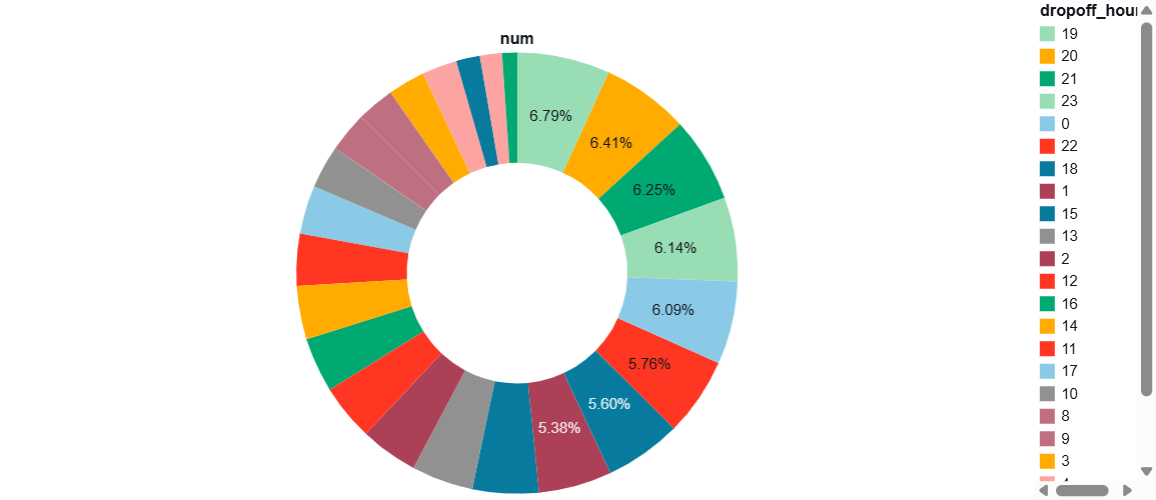
**Scatter**

****

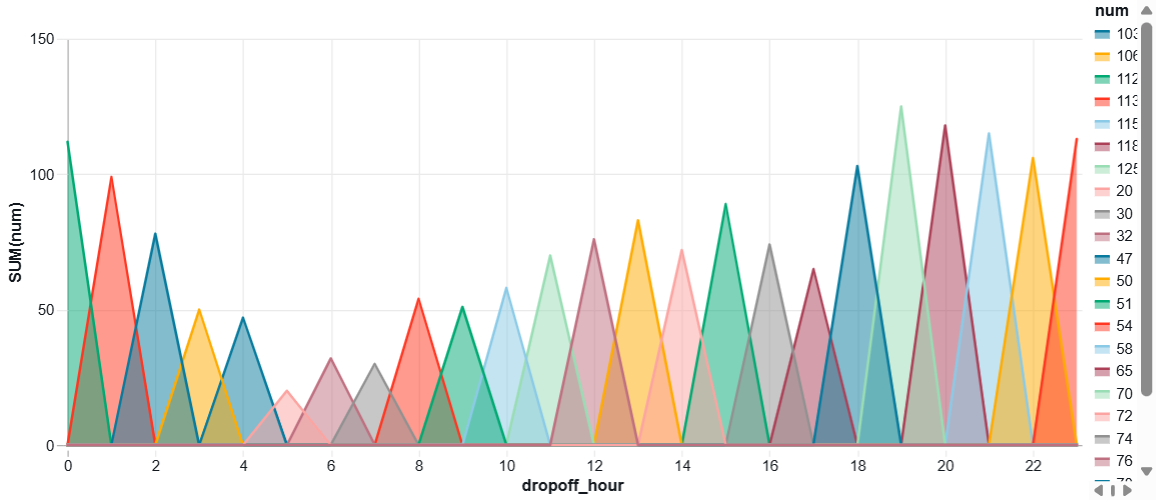
**Line**

****

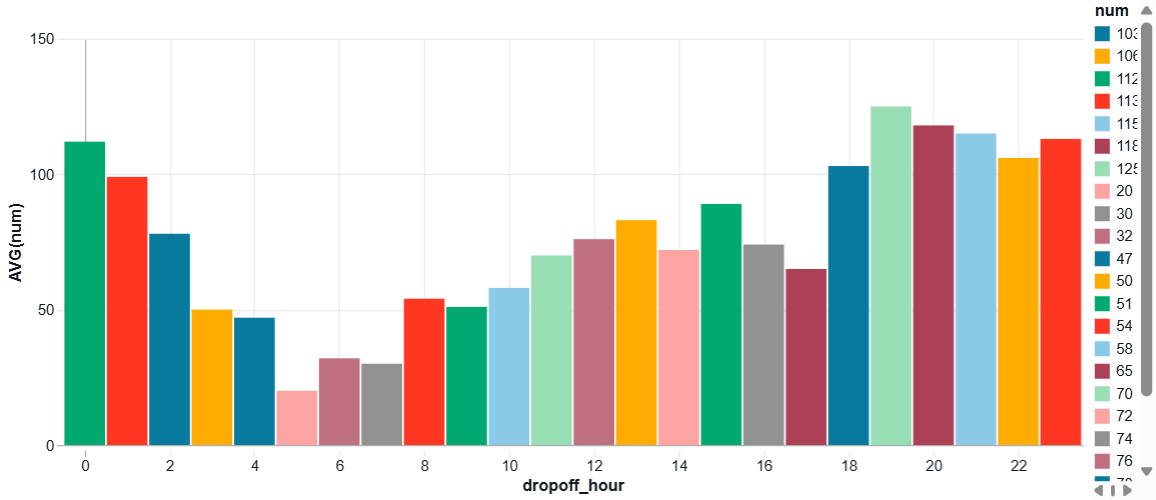
**Bubble**

****

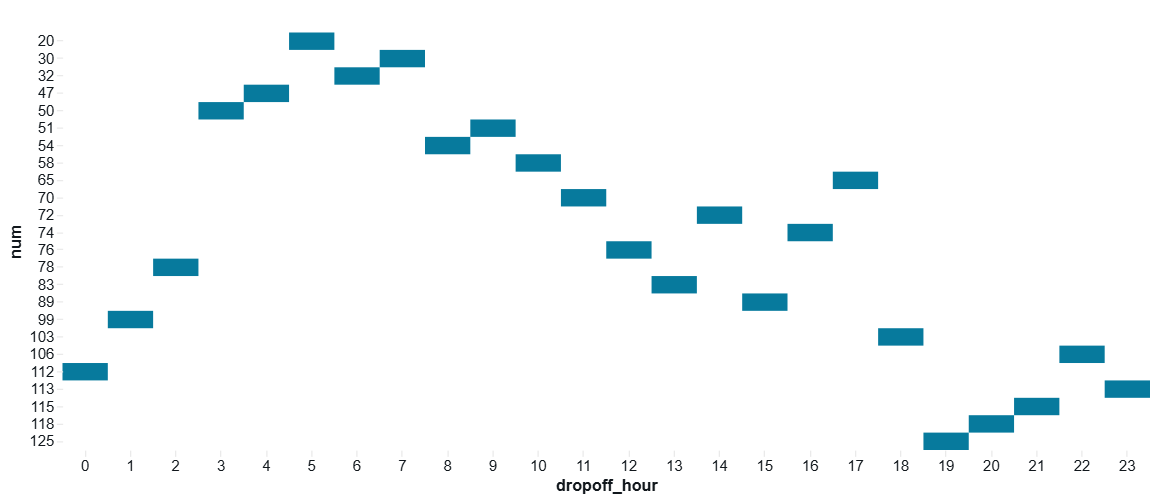
**Pie**

****

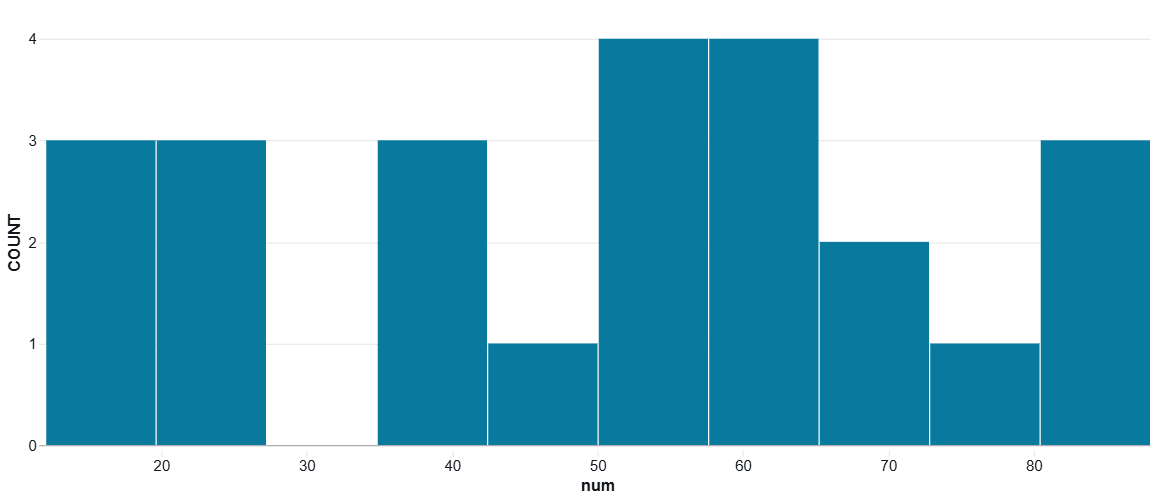
**Area**

****

**Bar**

****

**Heatmap**



**Histogram**

**%Python**

from pyspark.sql.functions import hour, col

pickupzip = '10001'  # Example value for pickupzip

df = spark.table("samples.nyctaxi.trips")

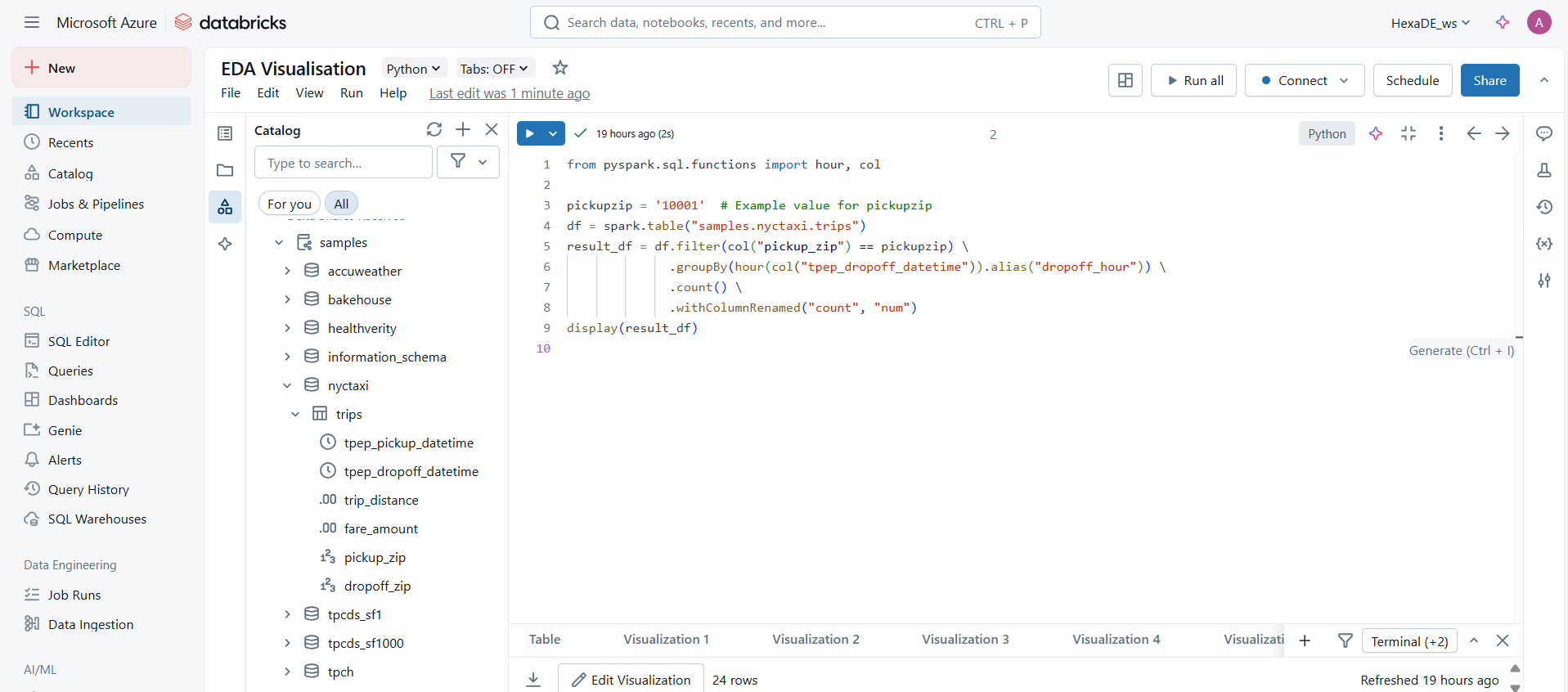
result\_df = df.filter(col("pickup\_zip") == pickupzip) \

              .groupBy(hour(col("tpep\_dropoff\_datetime")).alias("dropoff\_hour")) \

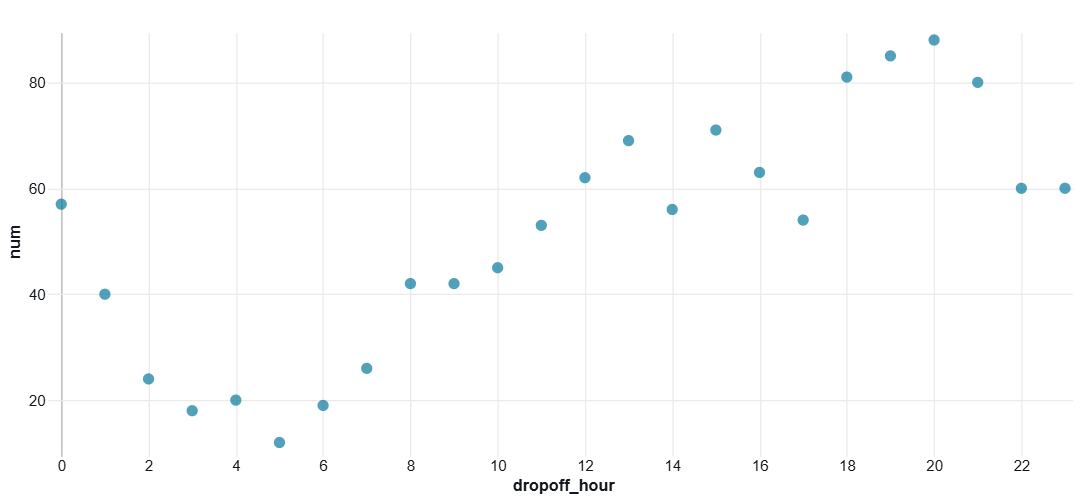
              .count() \

              .withColumnRenamed("count", "num")

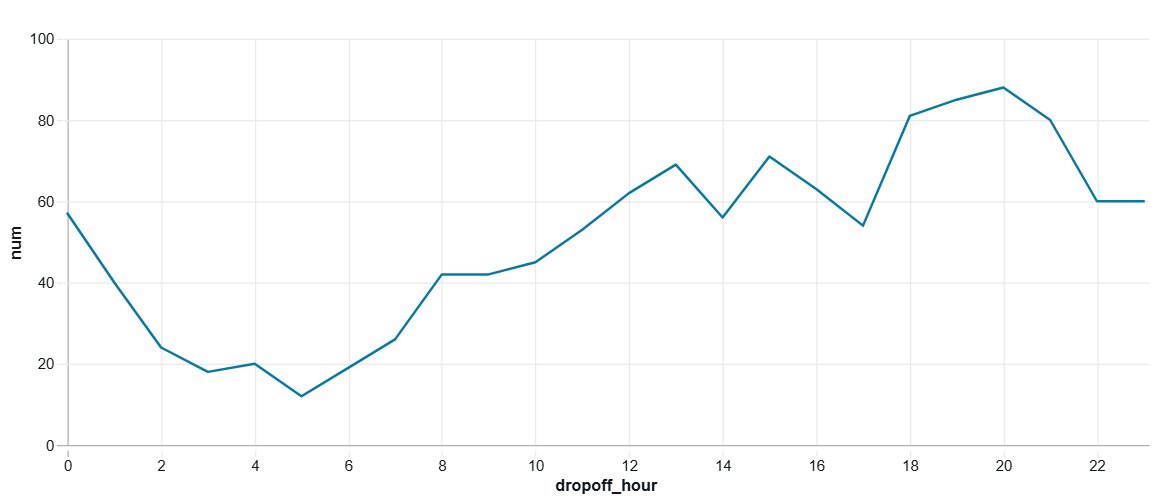
display(result\_df)

****

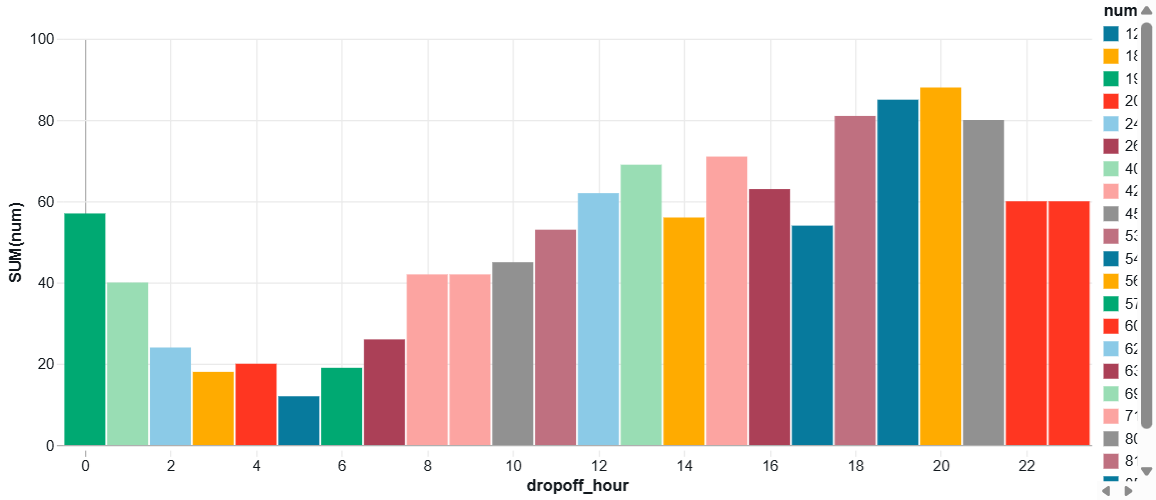
**Visualisations Types**

****

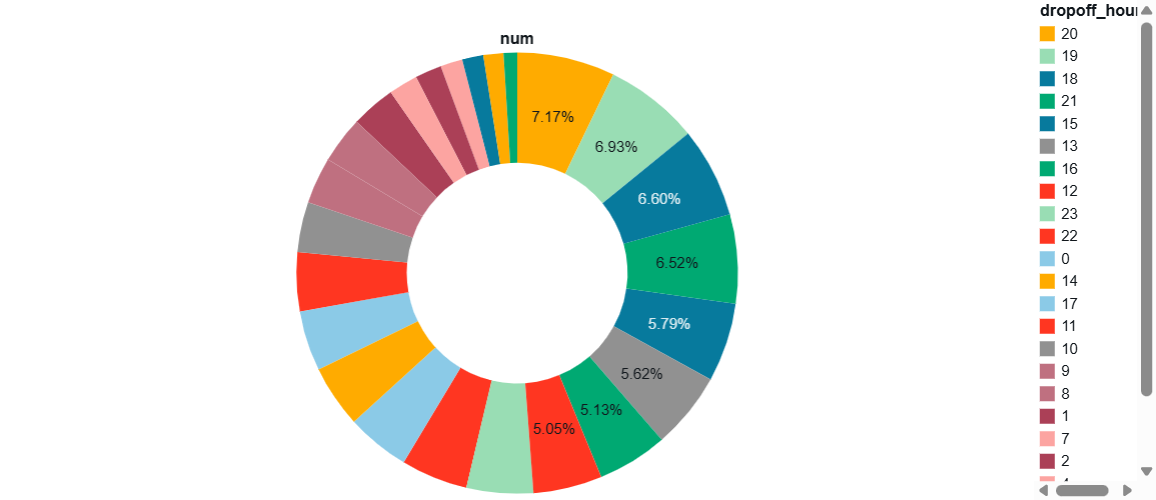
**Scatter**



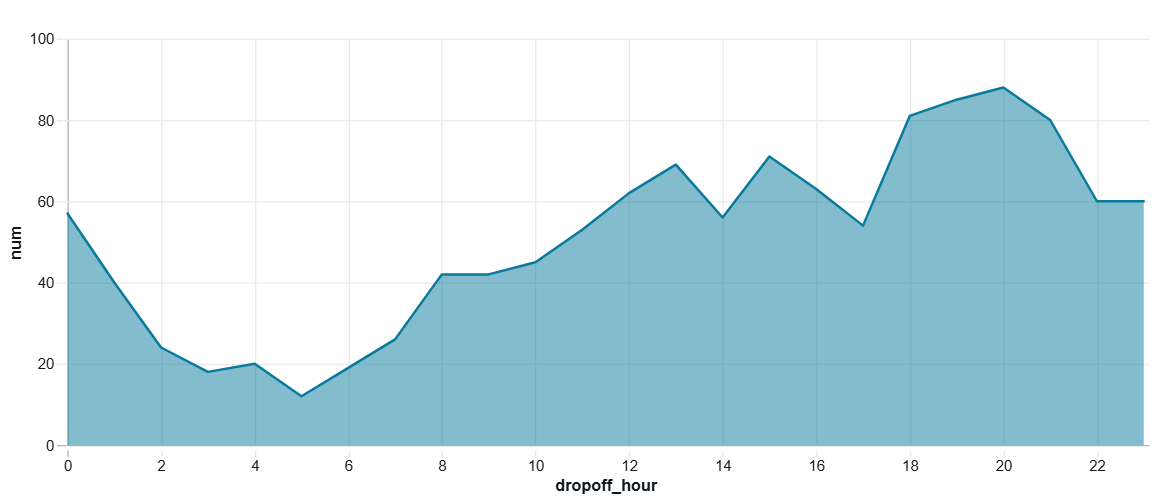
**Line**



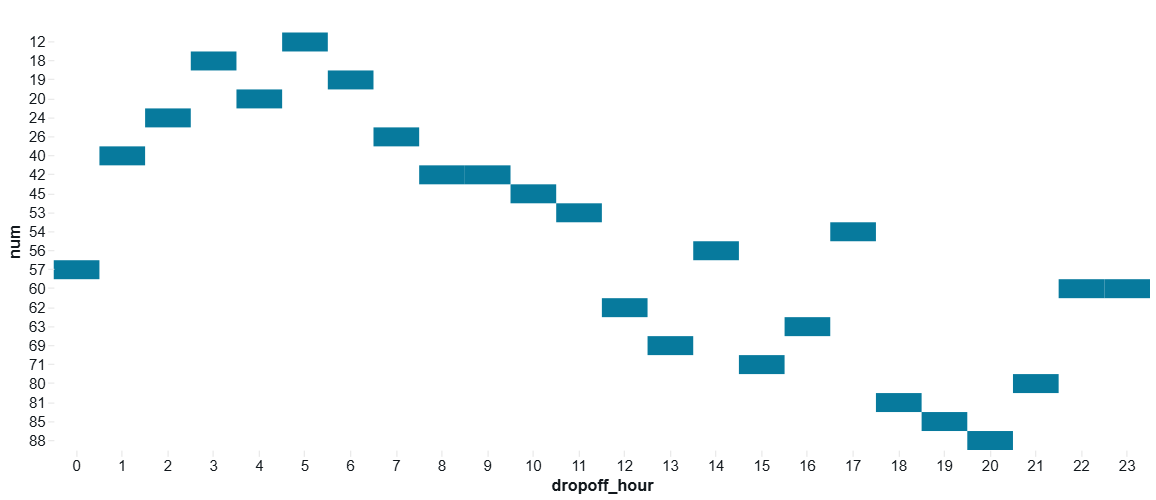
**Bar**



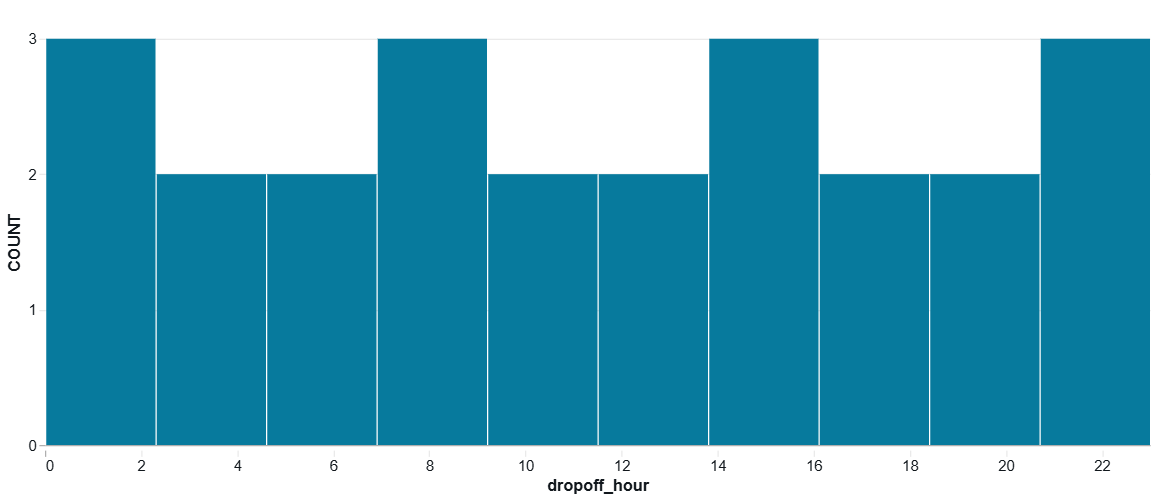
**Pie**



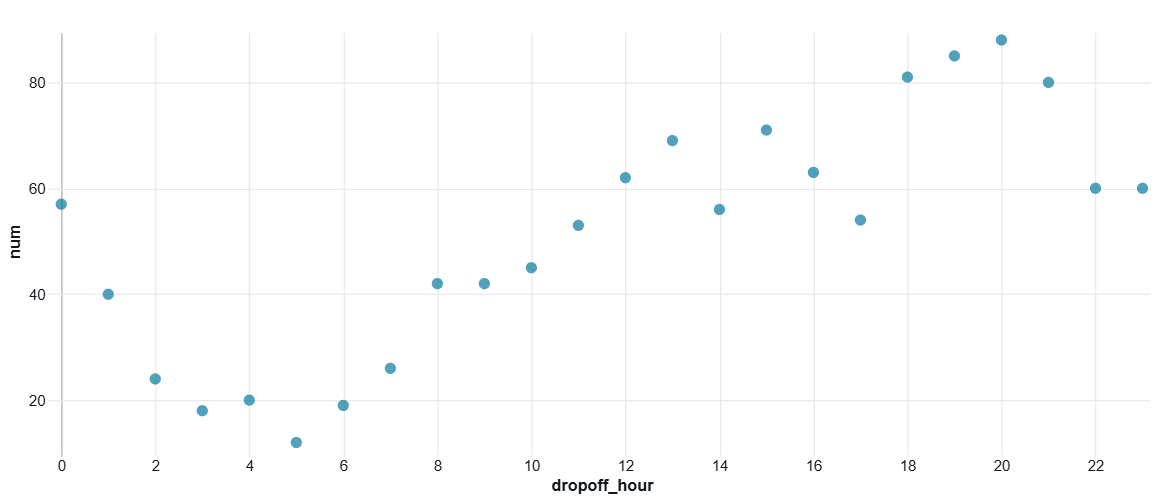
**Area**



**Heatmap**



**Histogram**

****

**Bubble**

**Submitted by**

Siva Balan T