

EDA VISUALIZATION

Building Queries and Visualizations

SQL

USE CATALOG samples;

SELECT

hour(tpep_dropoff_datetime) as dropoff_hour,

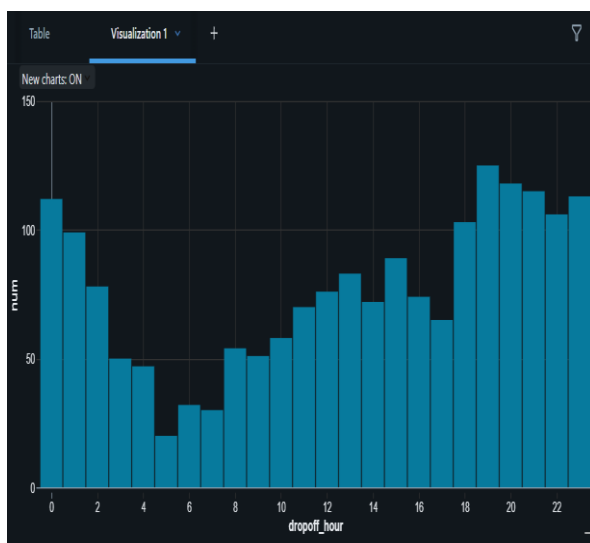
COUNT(*) AS num

FROM samples.nyctaxi.trips

WHERE pickup_zip IN ('10001', '10002')

GROUP BY 1;

Table	Visualization 1	+
1 ² ₃ dropoff_hour	1 ² ₃ num	
1	89	
2	118	
3	115	
4	106	
5	65	
6	99	
7	113	
8	47	
9	83	
10	54	
11	32	
12	50	
13	74	
14	70	
15	112	



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)
```

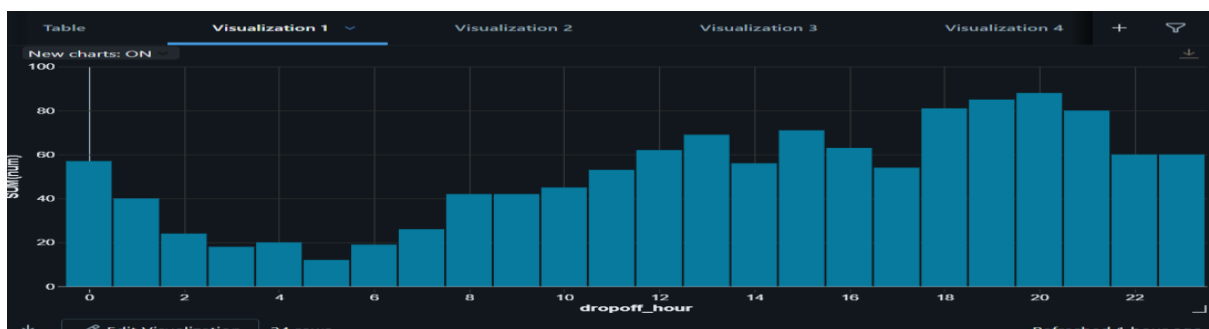
TABLE OUTPUT:

	¹² ₃ dropoff_hour	¹² ₃ num
1	15	71
2	20	88
3	21	80
4	22	60
5	17	54
6	1	40
7	23	60
8	4	20
9	13	69
10	8	42
11	6	19
12	3	18
13	16	63
14	11	53
15	0	57

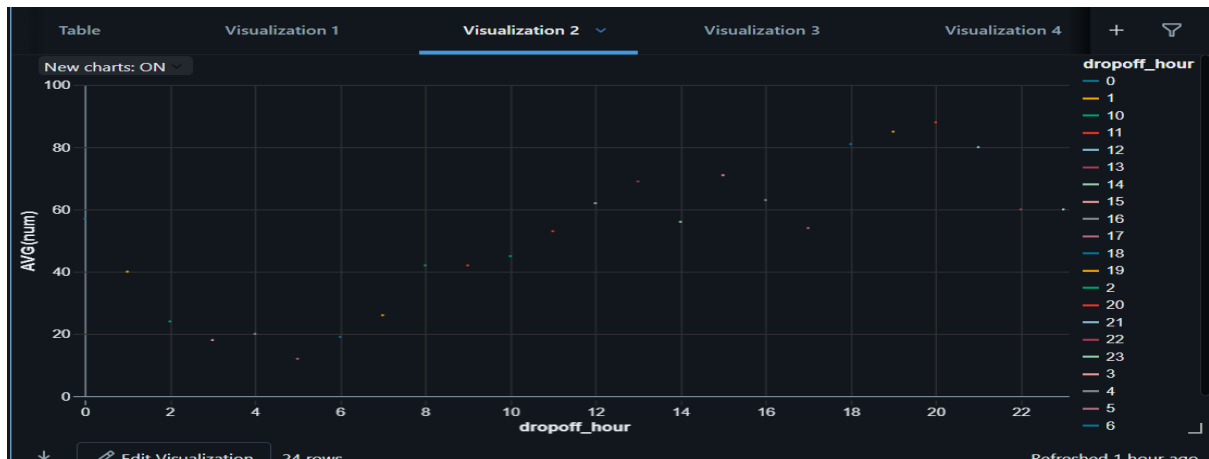
↓ ⌵ 24 rows | 3.67s runtime

VISUALIZATION OUTPUT:

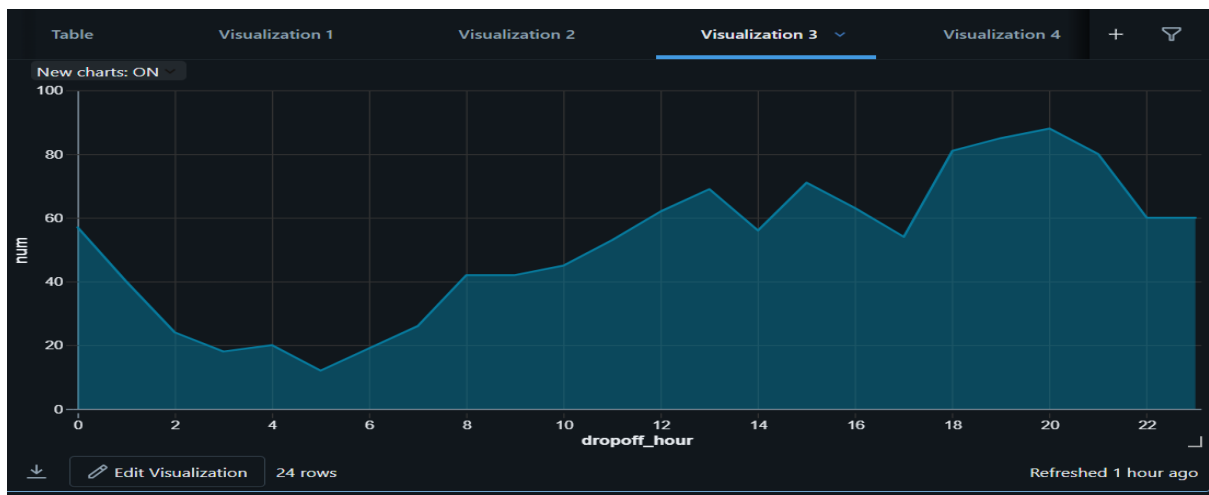
BAR CHART



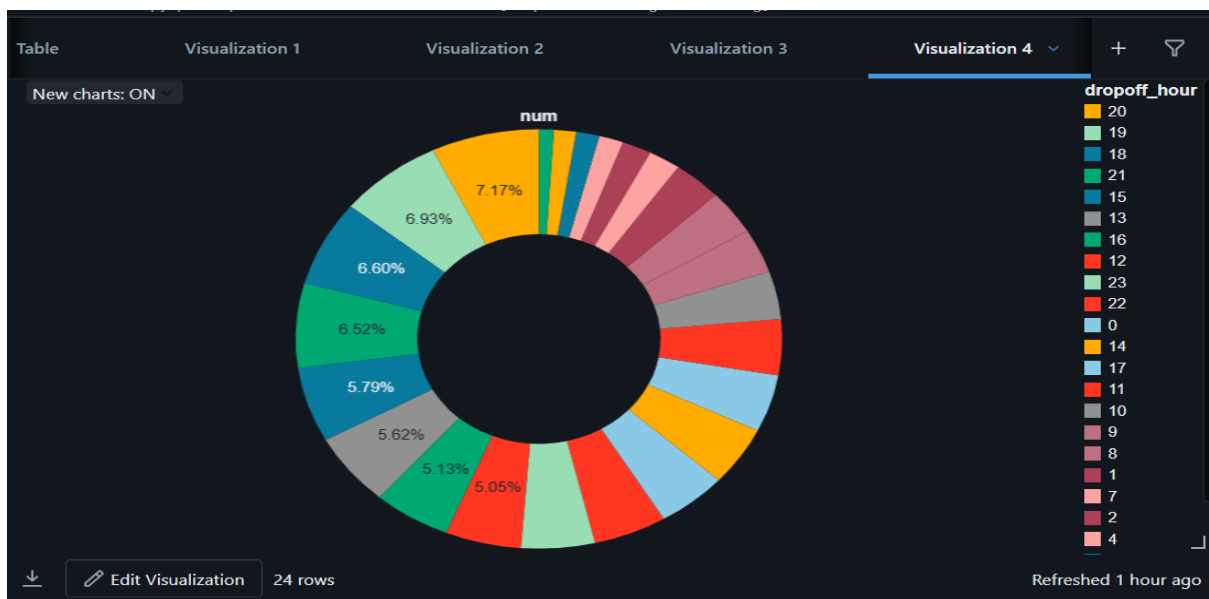
LINE CHART



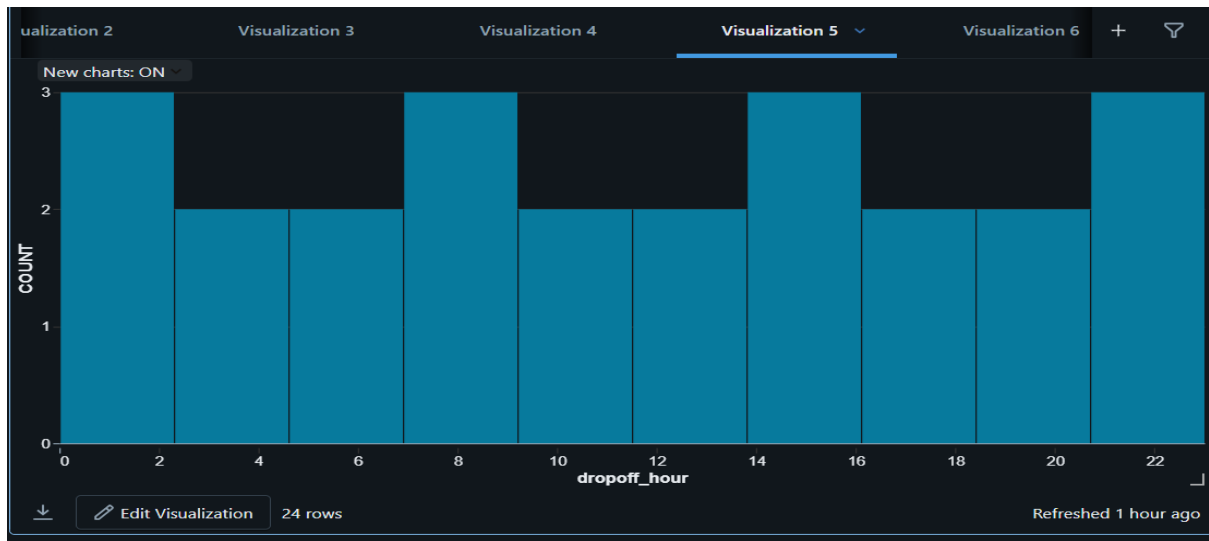
AREA CHART



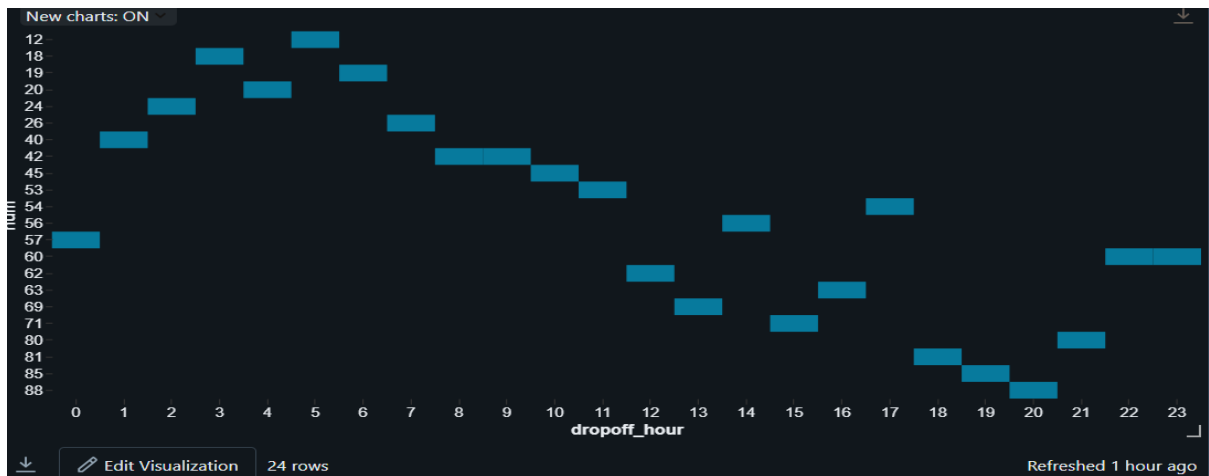
PIE CHART



HISTOGRAM



HEAT MAP



BUBLE CHART

