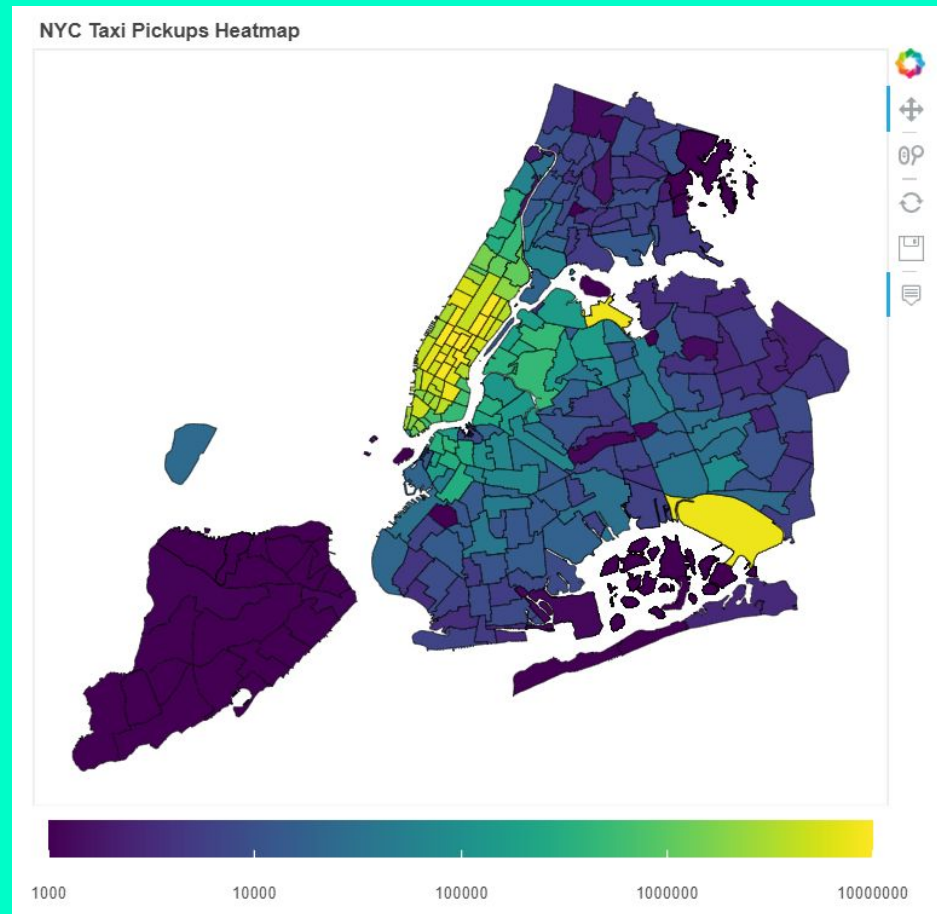
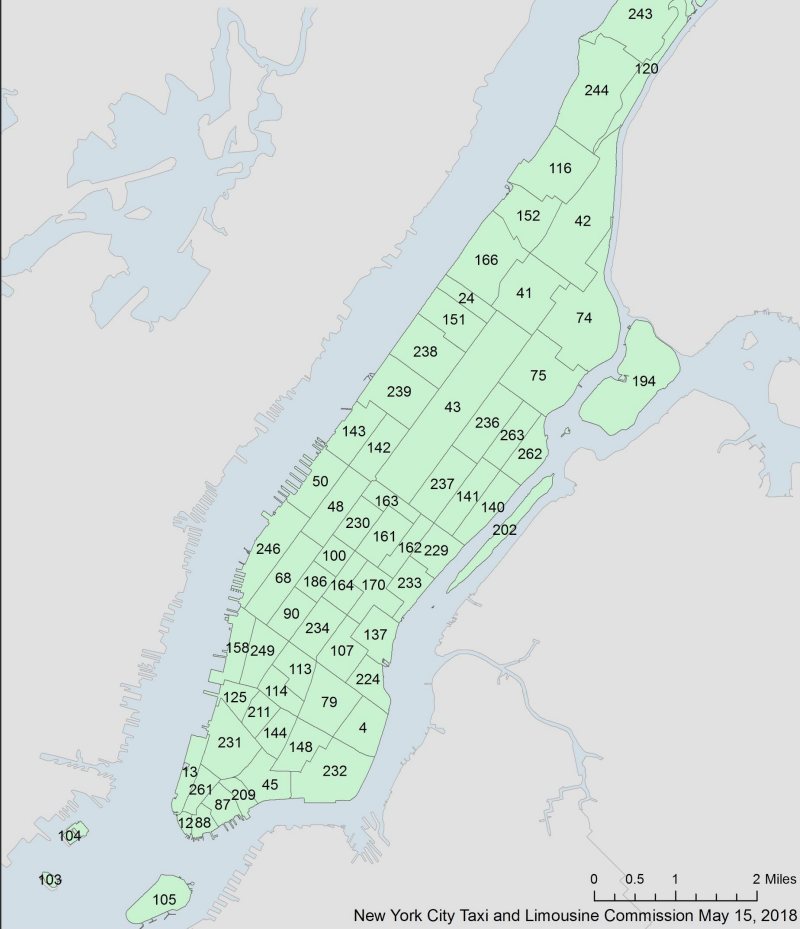


# NEW YORK TAXI RIDES

an area-based time-series  
prediction



# TLC Taxi Zones Manhattan



DATA

**freely available on:**

<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

**since 2017**

coordinates of pickups and dropoffs  
got replaced by zones

**data shows every NYC taxi ride**

data for 2017-2019 is around 25GB

=> transformation to time-series  
bins

# DATA TRANSFORMATION

VendorID

**tpep\_pickup\_datetime**

**tpep\_dropoff\_datetime**

passenger\_count, trip\_distance

RatecodeID, store\_and\_fwd\_flag

**PULocationID, DOLocationID,**

payment\_type, fare\_amount,

extra, mta\_tax, tip\_amount, tolls\_amount,

improvement\_surcharge, total\_amount

time\_bin

2017-01-01 00:00:00

2017-01-01 00:30:00

2017-01-01 01:00:00

...

rides

10

15

18

...

for each zones (265 total)

for each pickup and dropoff

calendar features, lags, weather

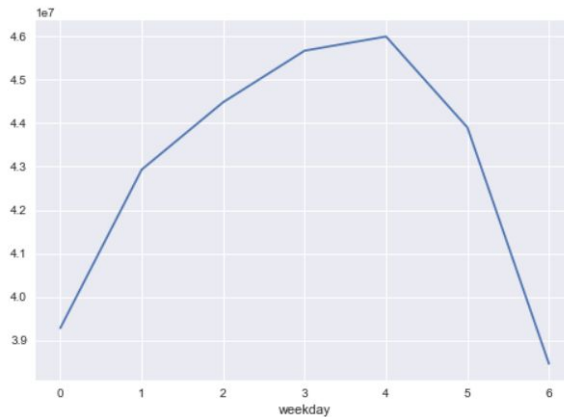
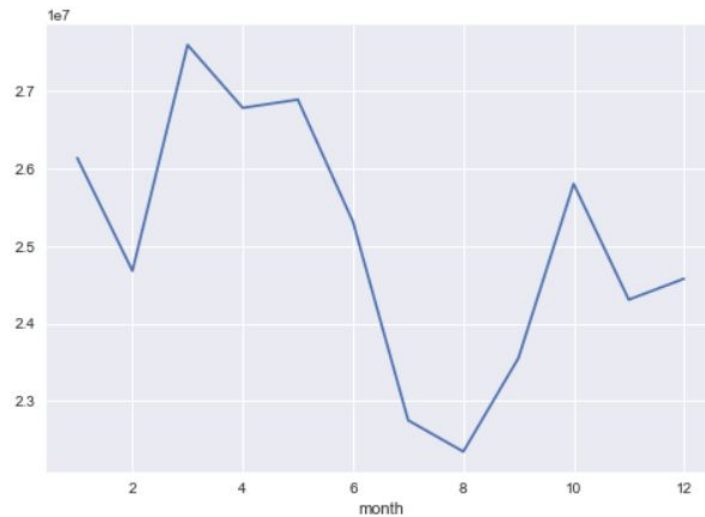
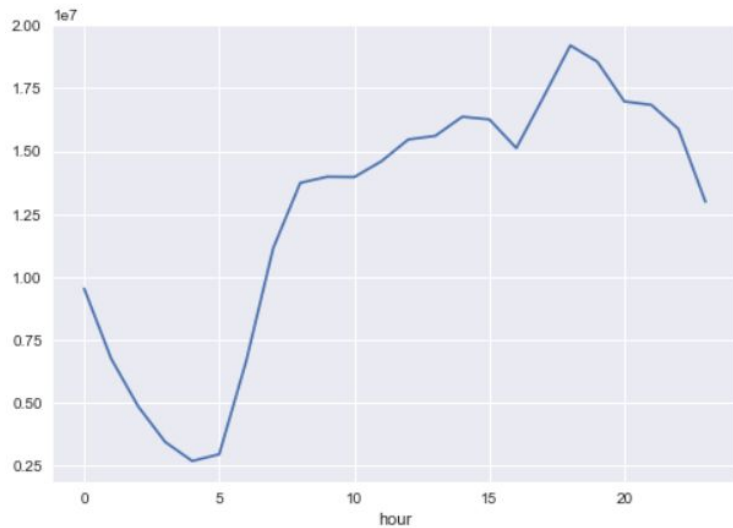
year, month, hour, weekday, holiday, tem

p, feels\_like, temp\_min, temp\_max, humi

dity, wind\_speed, weather\_main

# EDA

```
rides      1.000000  
lag168     0.933082  
lag1       0.924502  
lag24      0.864511  
lag167     0.863745
```



# RIDE DISTRIBUTION