DSH DISPATCH DATA ANALYSIS

Project Proposal

Gehad Ahmed

DATA DESCRIPTION

Datasets provided:

1. courier_wave_info_meituan.csv
   Contains details about courier waves and orders.
2. dispatch_rider_meituan.csv
   Includes dispatch details for riders.
3. dispatch_waybill_meituan.csv
   Contains dispatch waybill information.
4. all_waybill_info_meituan_0322.csv
   Holds comprehensive orders data.

DATA CLEANING

- Convert **date** from integer format (YYYYMMDD) to **datetime**
- Convert **Unix** timestamps to **datetime** format
- Convert **missing** timestamps to NaT (Not a Time)
- Convert **order_ids** in courier data from a string representation of a list to an actual list
- Fix the mismatch in **wave start time** being the courier data and orders data

KEY OBSERVATION

The maximum wave duration is **~12 hours**, which indicate an issue in the data.

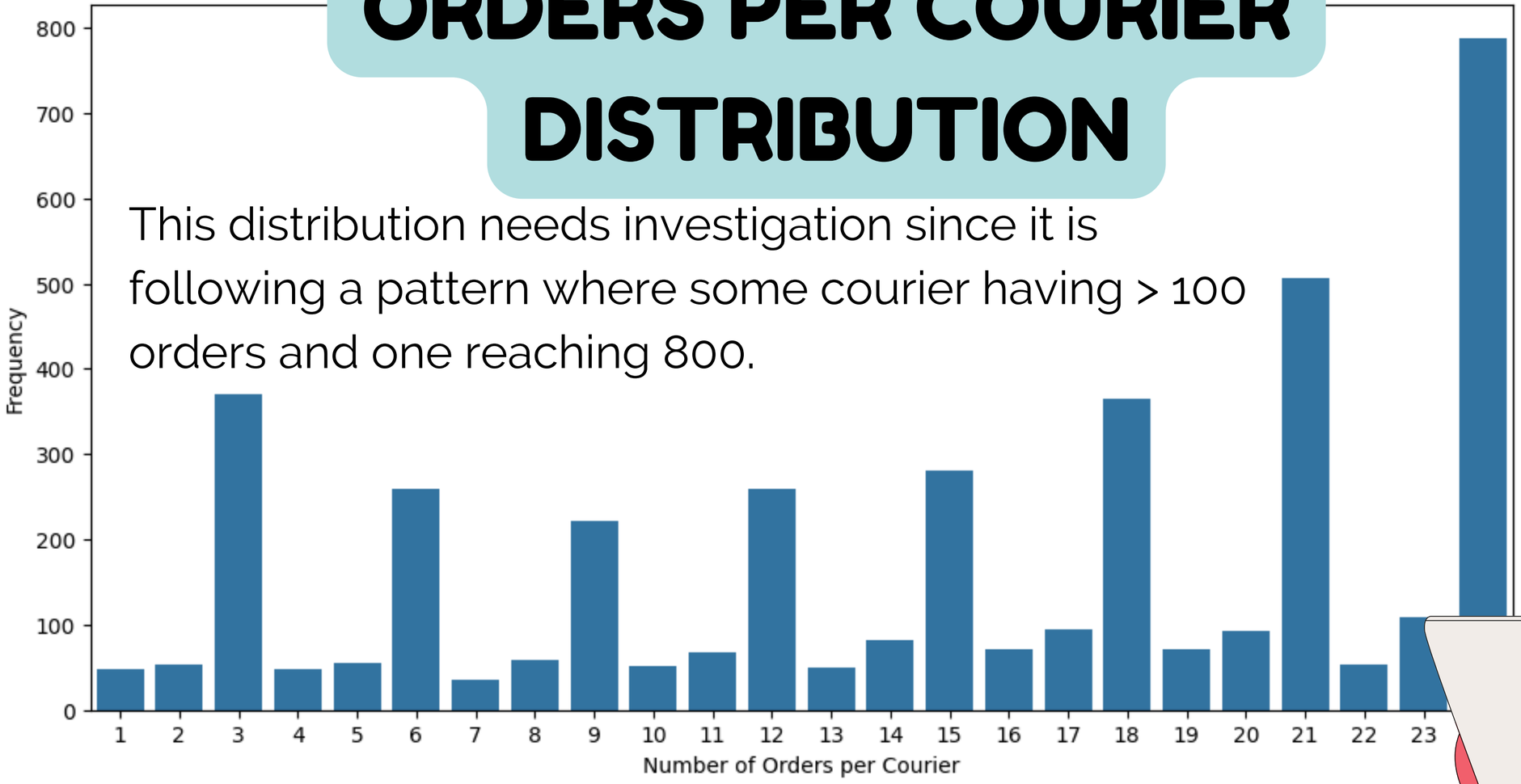
After investigation, the problem is with **prebooked orders** as they have their estimated meal time to be missing and hence they get dispached just after being ordered.

KEY OBSERVATION

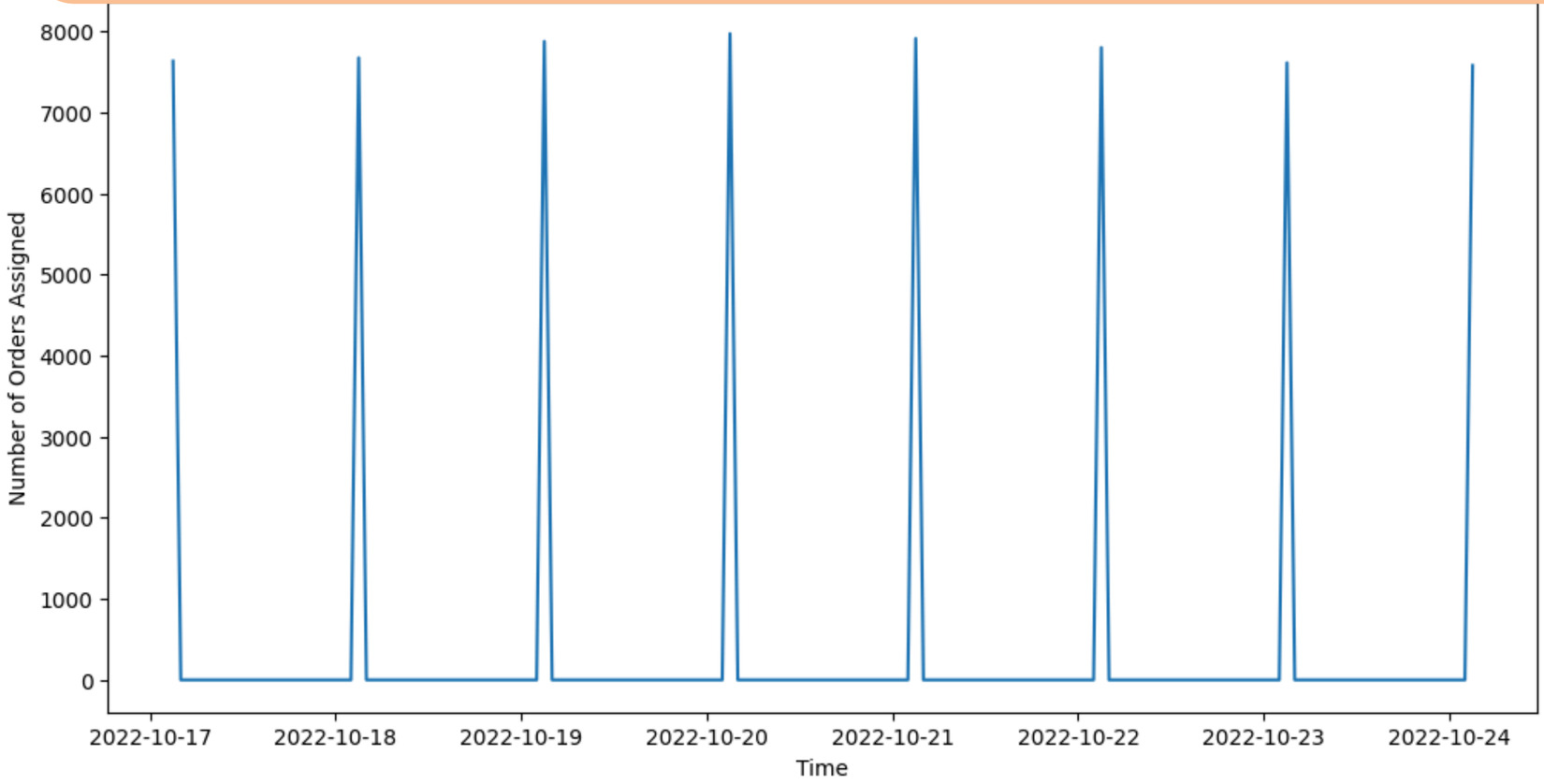
Metric	Prebooked Orders (24,109)	Non-Prebooked Orders (630,234)
Mean Time Gap	**891.86 sec** (~14.9 min)	**520.97 sec** (~8.7 min)
Standard Deviation	877.98 sec (~14.6 min)	392.44 sec (~6.5 min)
Minimum Time Gap	0 sec	0 sec
25th Percentile (Q1)	418 sec (~7 min)	241 sec (~4 min)
Median (Q2, 50%)	786 sec (~13 min)	472 sec (~7.9 min)
75th Percentile (Q3)	1,235 sec (~20.5 min)	745 sec (~12.4 min)
Maximum Time Gap	**36,897 sec** (~10.25 hrs) 🚨	**5,790 sec** (~1.6 hrs)

ORDERS PER COURIER DISTRIBUTION

This distribution needs investigation since it is following a pattern where some courier having > 100 orders and one reaching 800.



DISPATCH TIMES FOR COURIER



Ploting numbers of orders per hour show an issue with dispatch times in dispatch_rider.

Dispatches always happend **at Fixed Times** (03:25, 03:27, 03:30).

THANK YOU!