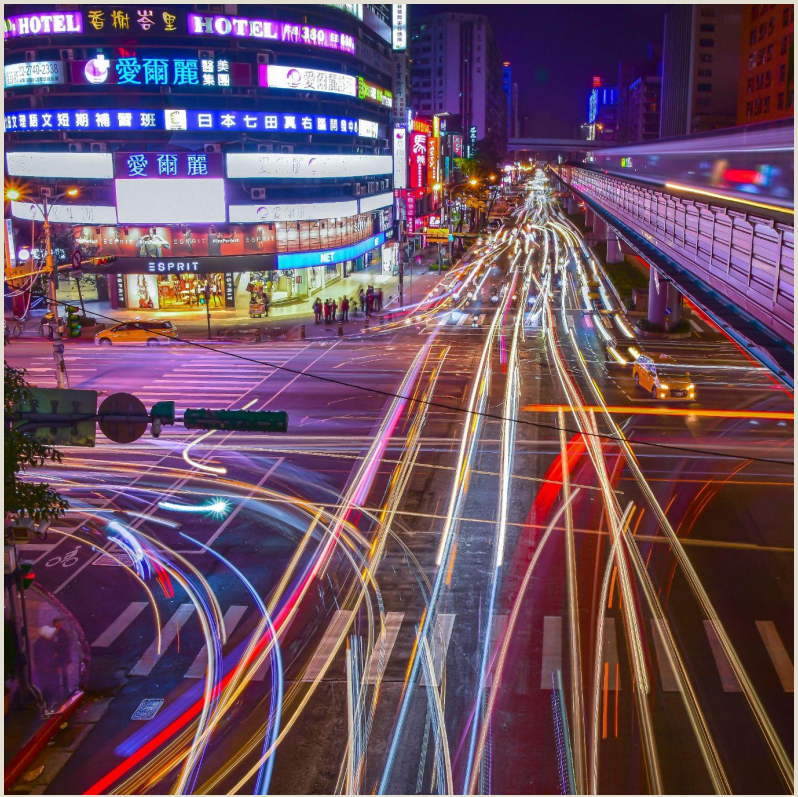
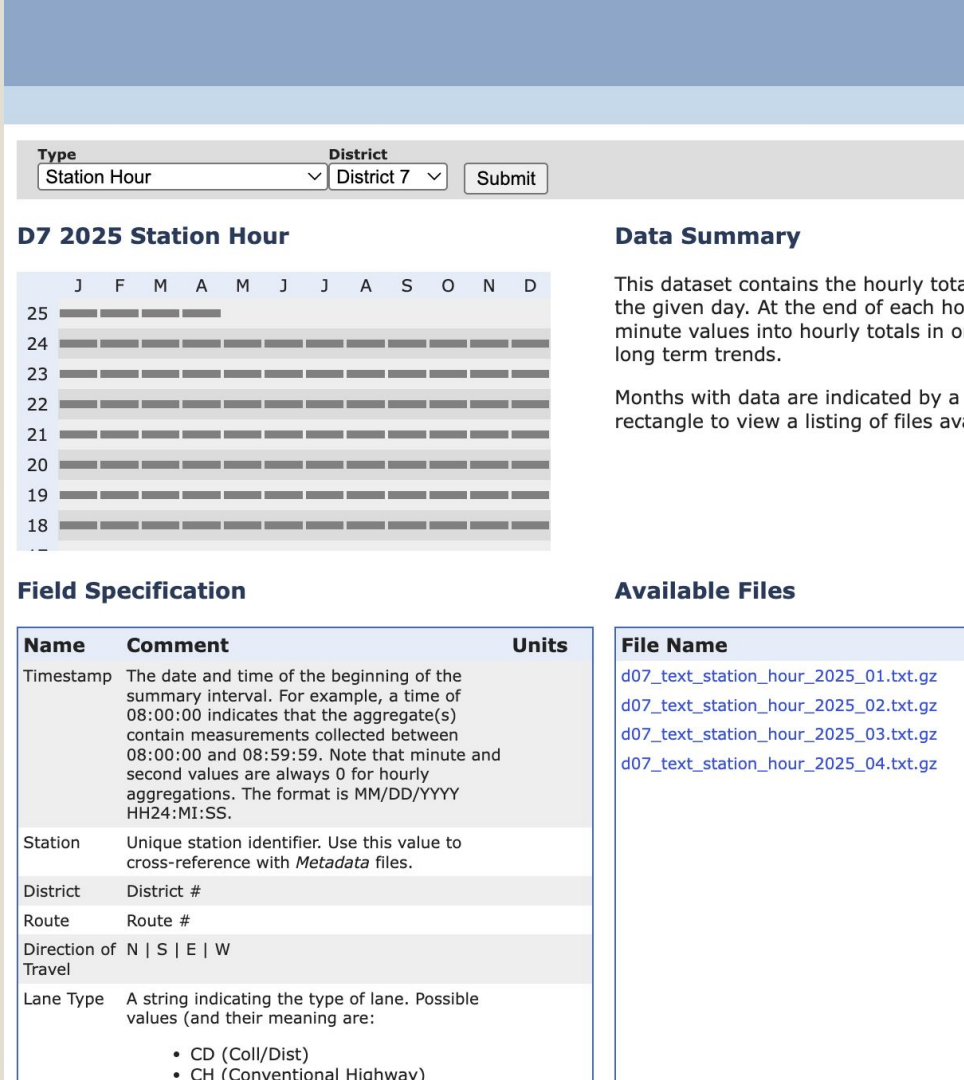
Traffic Prediction



LUCY LENNEMANN

1 GETTING THE DATASET

- Source: Caltrans Performance Management System
- Includes 10+ years of historical traffic data, i.e. lane closures, incidents, traffic counts
- Previously tried using the requests library
- Now using Selenium to webscrape



Type: Station Hour | District: District 7 | Submit

D7 2025 Station Hour

J F M A M J J A S O N D
25
24
23
22
21
20
19
18

Data Summary

This dataset contains the hourly tota
the given day. At the end of each ho
minute values into hourly totals in o
long term trends.

Months with data are indicated by a
rectangle to view a listing of files ava

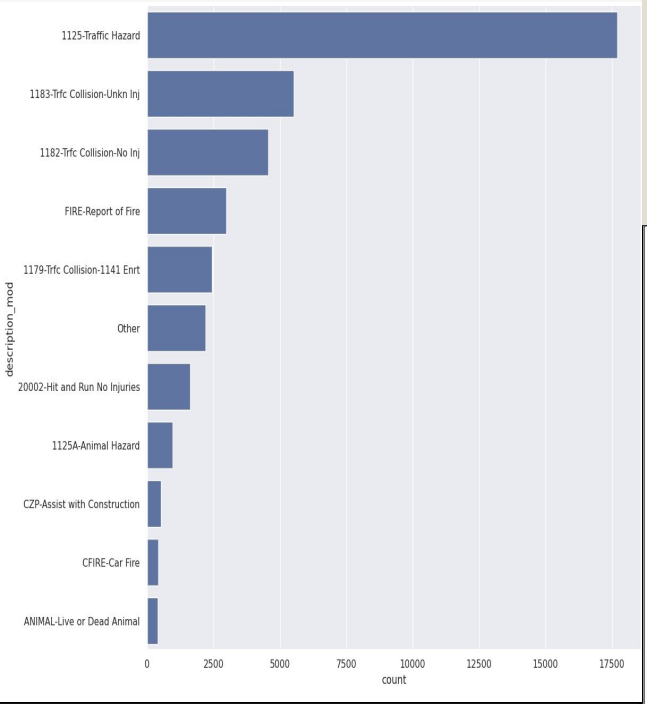
Field Specification

Name	Comment	Units
Timestamp	The date and time of the beginning of the summary interval. For example, a time of 08:00:00 indicates that the aggregate(s) contain measurements collected between 08:00:00 and 08:59:59. Note that minute and second values are always 0 for hourly aggregations. The format is MM/DD/YYYY HH24:MI:SS.	
Station	Unique station identifier. Use this value to cross-reference with *Metadata* files.	
District	District #	
Route	Route #	
Direction of Travel	N | S | E | W	
Lane Type	A string indicating the type of lane. Possible values (and their meaning are: • CD (Coll/Dist) • CH (Conventional Highway)	

Available Files

File Name
d07_text_station_hour_2025_01.txt.gz
d07_text_station_hour_2025_02.txt.gz
d07_text_station_hour_2025_03.txt.gz
d07_text_station_hour_2025_04.txt.gz

2. EXPLORATORY DATA ANALYSIS

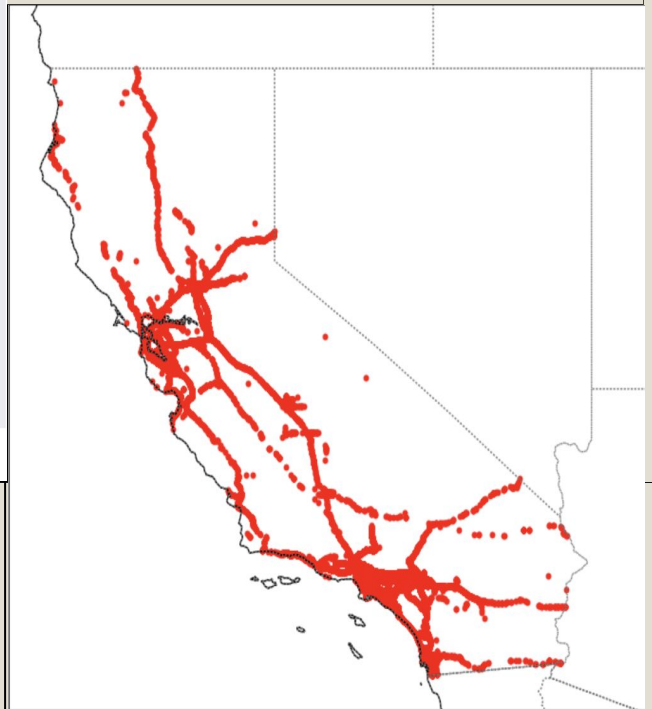


California Highway Patrol Incidents - Jan 2025

Top causes of incidents were **traffic hazards** and **traffic collisions.**

California Highway Patrol Incidents - Jan 2025

Locations of accidents correspond to major highways.



- Available datasets to scrape include CHP incidents, totals for each station, and more
- Narrow down useful datasets and fields
- **Messy data** with lots of missing values, unclear field names
- Probably a **large volume** of data; some of the datasets are aggregated monthly for over 10 years

Initial Insights

3 PROPOSED PRODUCT

- Allow users to enter 2 addresses and a date → predict amount of time it will take to drive from Location A to Location B
- Narrow down scope to LA
- Methodology: (1) get the rest of the data (2) clean and aggregate data (3) use a forecasting model (4) Streamlit for the dashboard