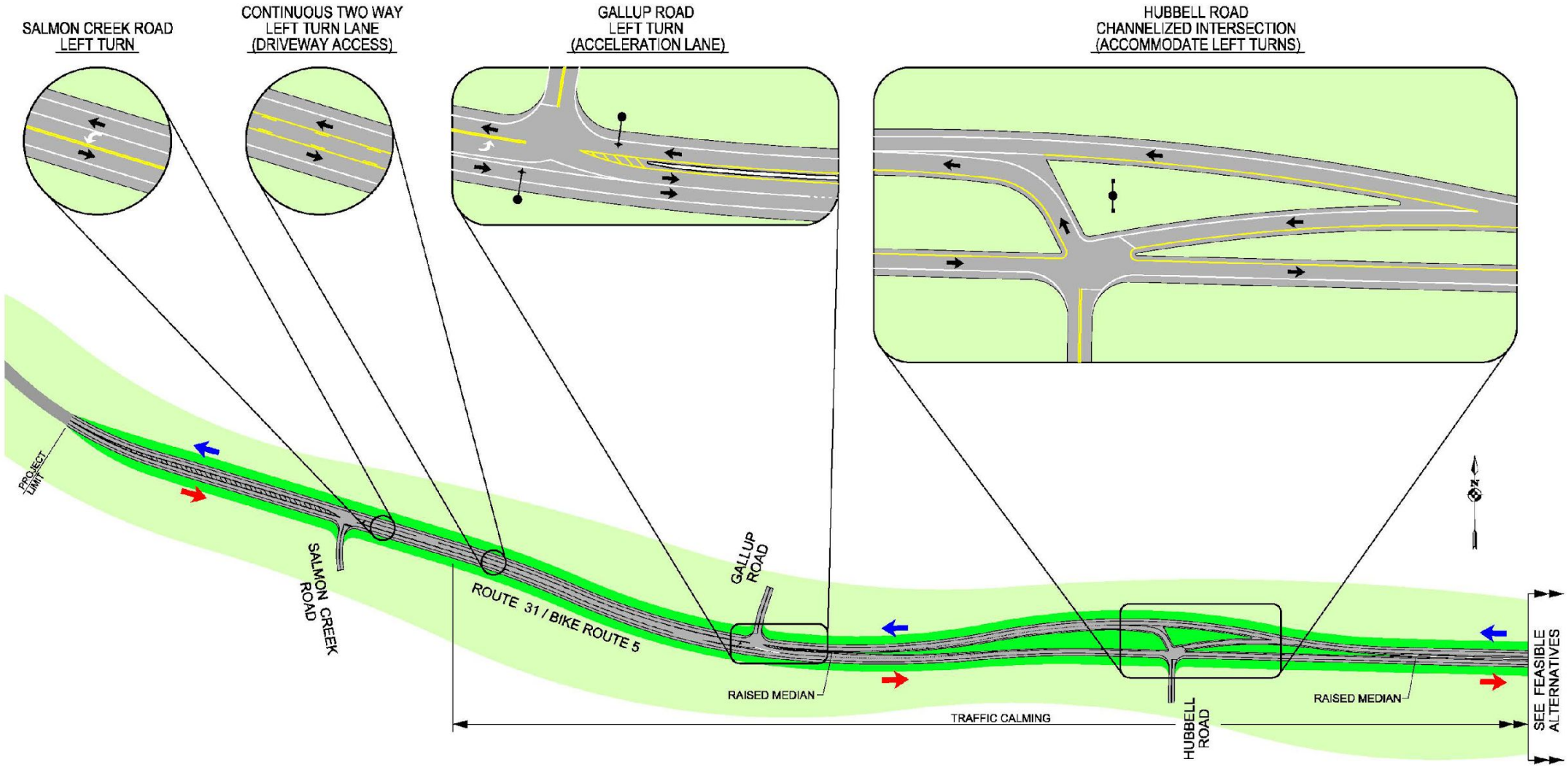
What is Data Science?

Digital Sensors everywhere

Two data analysis tasks

1. Establishing facts:
    1. The water in City **X** is safe/unsafe to drink.
    2. The average daily time spent on a smartphone is **Y**
    3. The number of almost accidents per-day on highway 5 is **Z**
2. Making predictions:
    1. Selecting ads to maximize clicks.
    2. Automating/assisting medical diagnostics.
    3. Choosing candidates for Loans.


CARPOOLS ONLY
2 OR MORE PERSONS
PER VEHICLE
CLICK IT
OR TICKET
Culver City
EXIT
49A



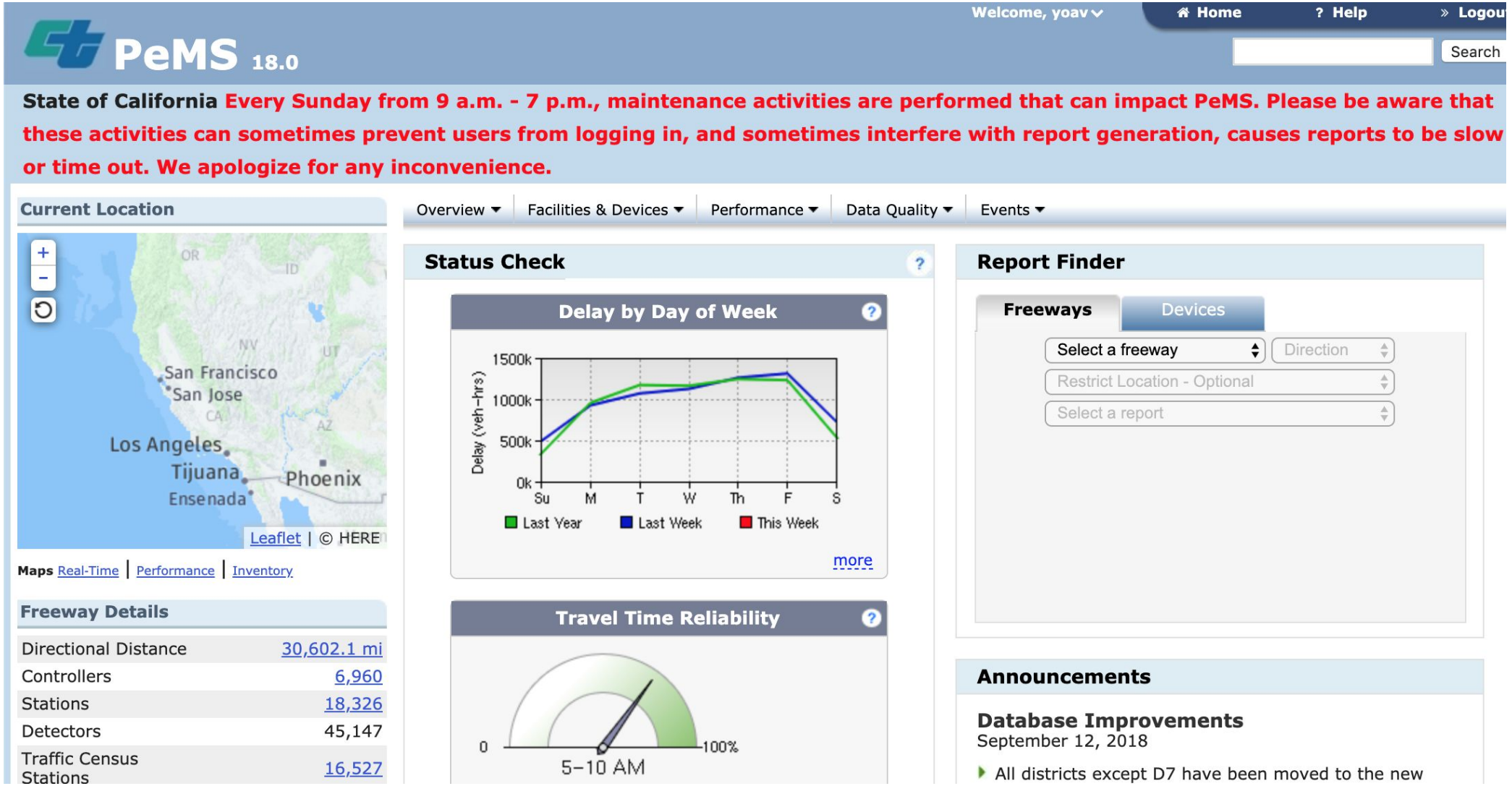
LEGEND

- RT 31 WESTBOUND
- RT 31 EASTBOUND
- POTENTIAL HIGH MAST LIGHTING
- POTENTIAL COBRA HEAD STREET LIGHTING

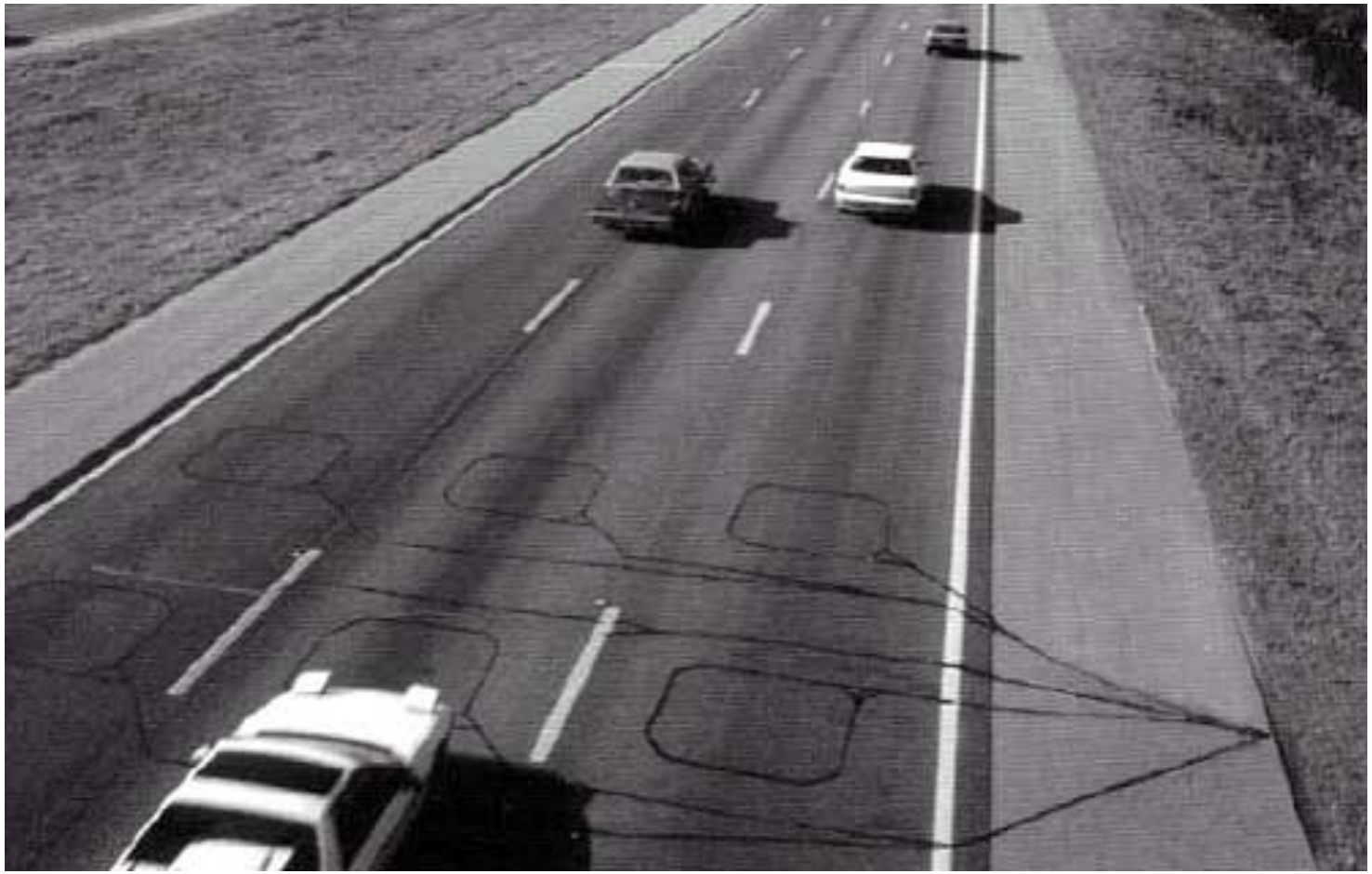

STOP
HERE
ONE
VEHICLE
PER
GREEN
GL BAL
Restaurant Equipment & Supplies
305-688-8700
EW & USED EQUIPMENT

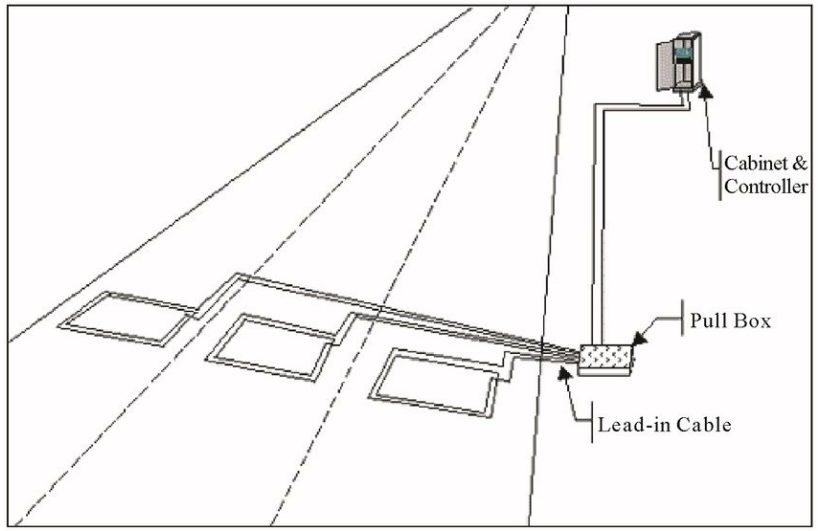
A real life example highway traffic in California.

Caltrans: PeMS (Performance measurement system)



Car Loop Detectors





- For each passing car:
  - **flow** : number of cars per unit time.
  - **Occupancy**: fraction of time that there is a car above the coil.
- ~45,000 individual detectors.
- Each detector generates a data packet every 30 seconds.

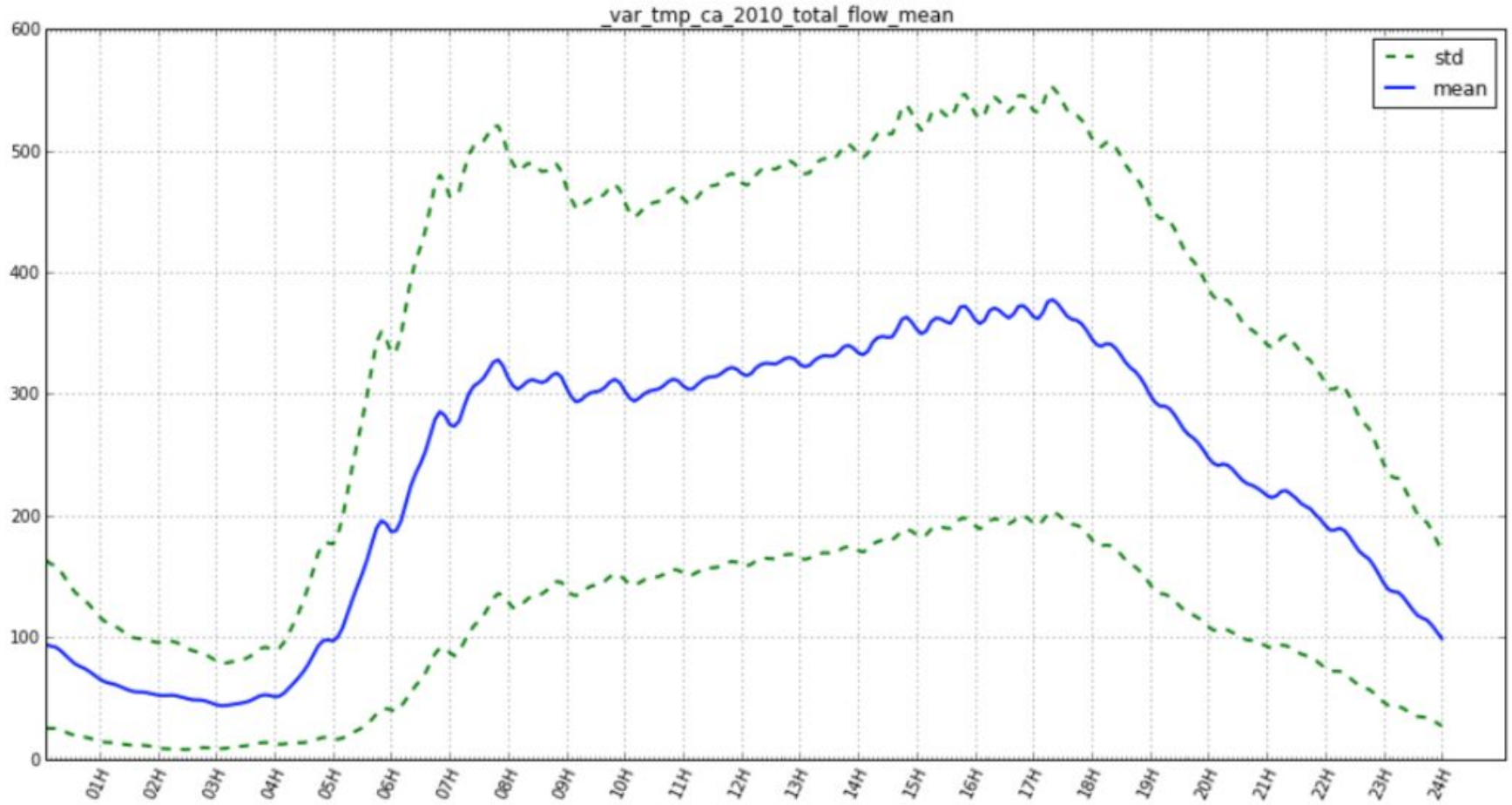
MAS-DSE Capstone Project, 2016

Kevin Dyer, John Gill III, Conway Wong

Caltrans PeMS

The primary data source chosen originates from the California Department of Transportation (CalTrans) Performance Measurement System (PeMS). The PeMS system is an Archived Data User Service (ADUS) site providing more than 10 years of historical traffic data collected from more than 39,000 sensors across California highways and major roadways. The site archives CalTrans loop detector data from traffic management centers (TMCs) throughout the state. The TMC's main purpose is to provide a portal which exposes analytical capabilities for various use cases such as supporting freeway operations, travel, and research. Additionally, it hosts a publically available data warehouse called the Data Clearinghouse. The Data Clearinghouse provides freely available historical data collected from sensors all across the state. The Data Clearinghouse was the source of the majority and most important data sets used for this Capstone project.

Mean and STD of total flow (cars per minute)



Top Eigen-vectors