
CSE 545

Good Health and Well Being

Bengaluru Bigdata Boys

Abhiram M Kaushik, Ajay Gopal Krishna, Rajat R Hande, Rajesh Prabhakar

Introduction

Give pointers to achieve the SDG Goal-3 related to

- Premature mortality
- Health impacts, caused by

Air pollution



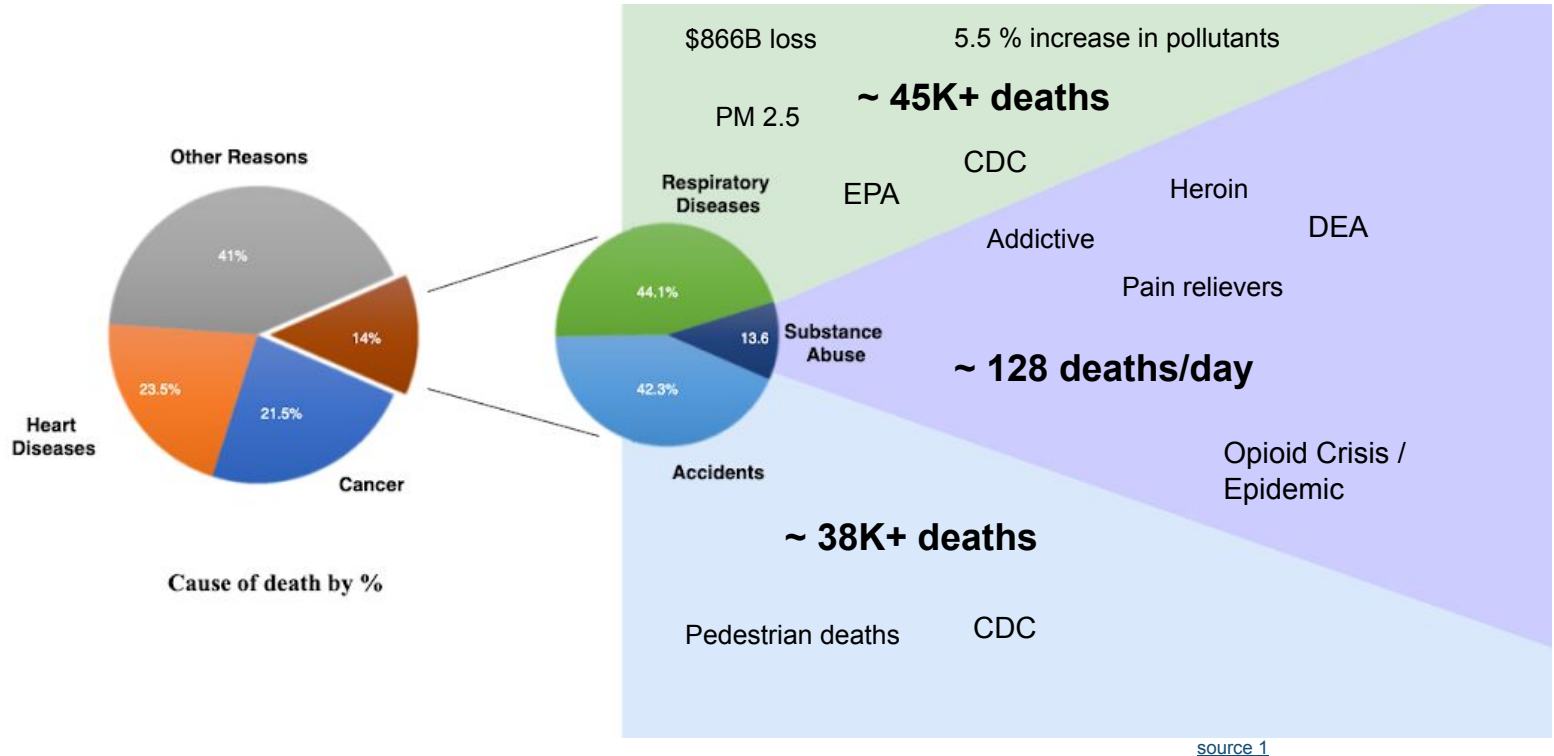
Substance abuse



Traffic Accidents



Motivation



Background

Related

- Health impacts due to Air pollution, 2019 [[paper](#)]
- Traffic accidents rates in the US [[paper](#)]
- Leading causes of deaths in US 2017(NVSS): [[paper](#)]
- California Air Pollution Analysis: [[paper](#)]

Research:

Methods

General Pipeline

Data (Pre)processing

- Data Cleaning
- Data Transformation
- Data Reduction
- Data Storage



Spark/Hadoop

Spark



Data Analysis

- Time Series Analysis
- Multivariate **Linear Regression**
- **Hypothesis Testing**

Results and Conclusions

- Result Analysis
- Data Visualization



Database/Python/JS



Methods

Mortality dataset from [CDC](#) (4GB). This needs to be cleaned and extracted into csv files.

Health Impacts due to Air Pollution

- Pollution dataset: [EPA](#) (1 GB), primarily based on county level AQI.
 - Perform Linear Regression in **Spark**
 - compute correlation of multiple pollutants against mortality rate in the US.
 - Use results to:
 - p-test (validation)
 - mortality prediction
-

Methods

Substance Abuse

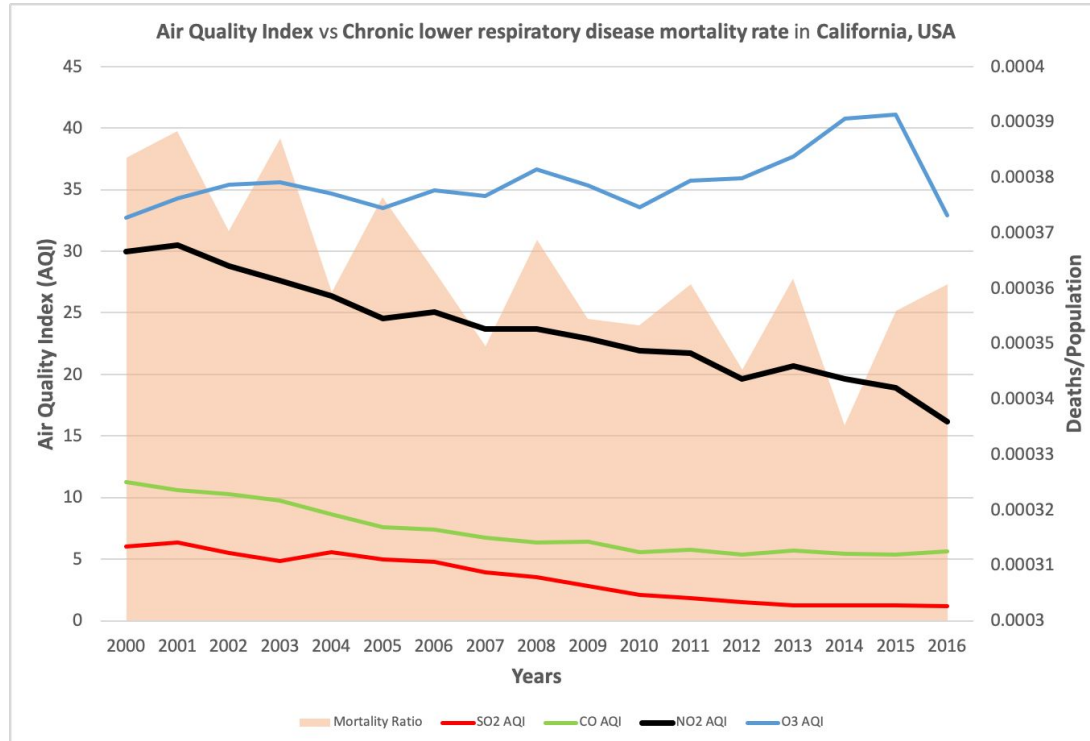
- Extract Opioids purchase data from [DEA](#).
- Perform clustering to detect addictive behaviour through buying patterns.

Traffic Accidents

- Traffic dataset from [here](#).
- Analyse the data to identify current and future hotspots.

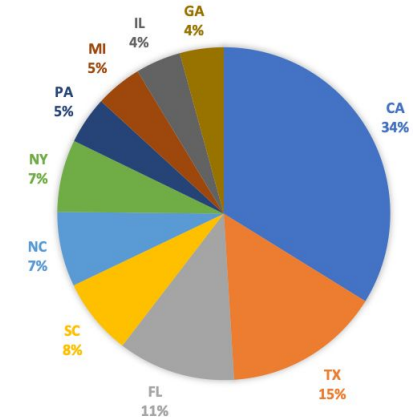
Validation of the above results through hypothesis testing.

Preliminary Results



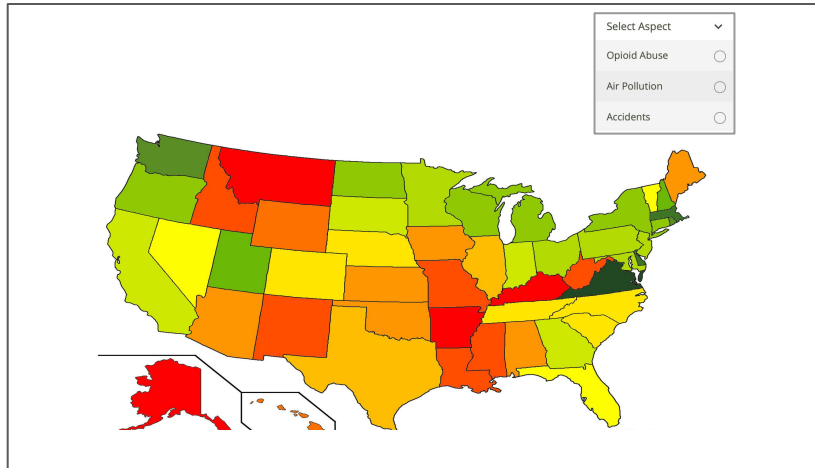
$\text{Corr}(\text{NO}_2 \text{ AQI, Mortality Rate}) = 0.7370$
 $p\text{-value} = 0.0007368$

STATE WISE ACCIDENTS BREAKDOWN



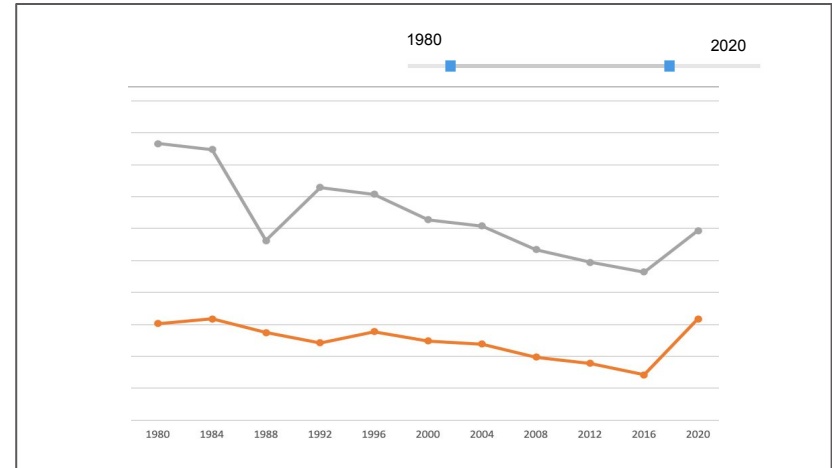
Expected Results

Heatmap showcasing the mortality rates in various states due to the different aspects



It can be further drilled down to the county level.

Time Series Analysis



Conclusion

After performing our analysis, we want to explain the observations and provide suggestions to achieve the SDGs.

- Premature deaths: Identify mortality rates due to air pollution, substance abuse and traffic and suggest remedies.
 - Strengthen the prevention and treatment of substance abuse.
 - Identify hotspots of Air Pollution and traffic accidents.
-

Thank you
