CrimeChronicleLA

“A Big Data Exploration of Los Angeles Crimes”

Team:
Ashwathi Subramanian
Pugazhventhan Kulandaivel

CSCI 5253 - Datacenter Scale Computing

Contents

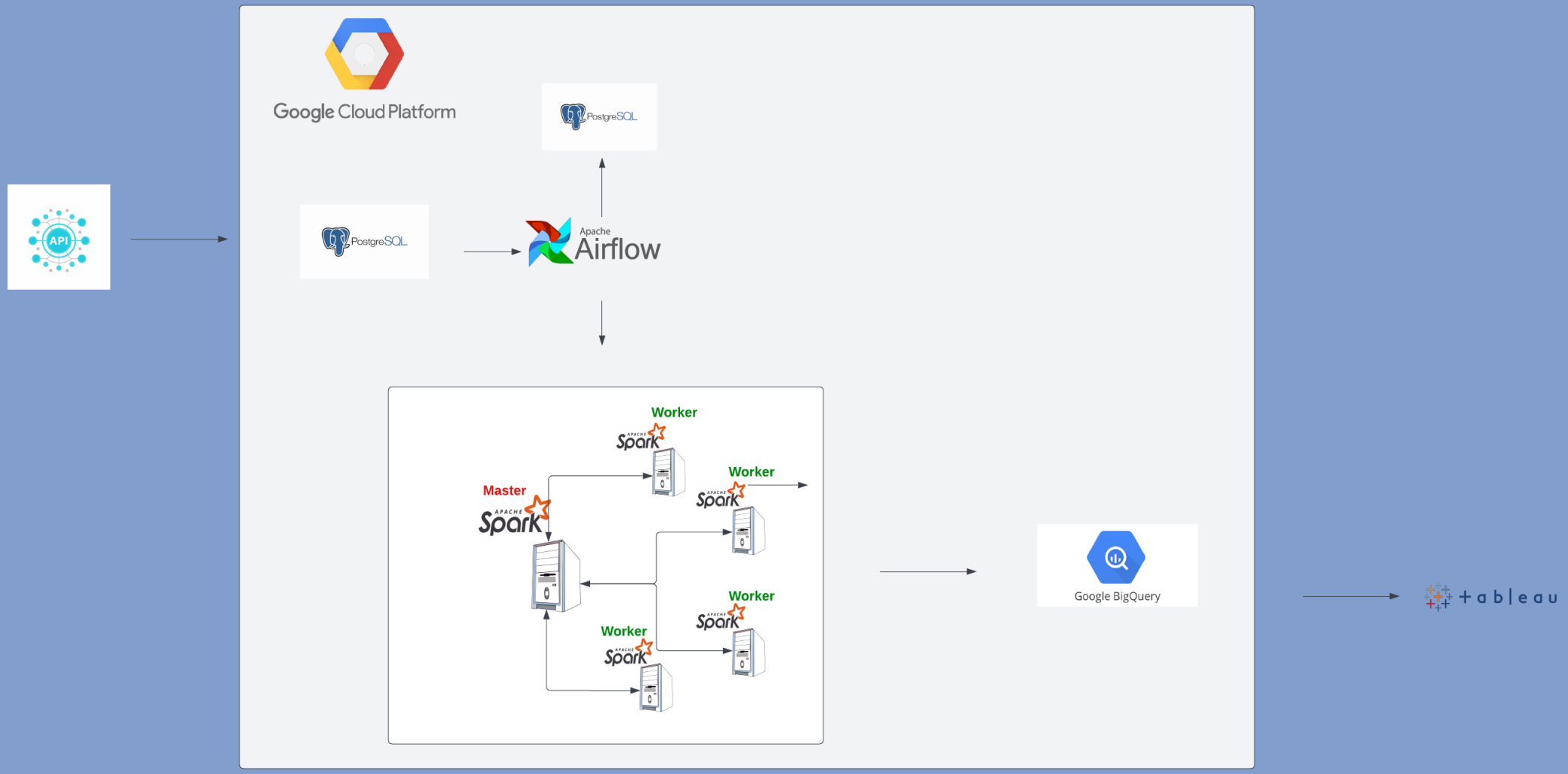
Introduction

This project focuses on analyzing Los Angeles crime data (2020-present) with objectives like **utilizing big data tools to process huge amounts of data**, in-depth data analysis, creating informative visualizations, and predicting crime rates. It aims to uncover patterns, identify common timeframes and crime types, and evaluate high-crime areas within Los Angeles Patrol Divisions.

Who benefits from this?

This project provides crucial insights into Los Angeles crime trends, benefiting law enforcement, urban planning, and public awareness. It supports resource allocation, strategic personnel deployment, and proactive measures, contributing to effective policing and empowering the community to make informed decisions for enhanced safety and quality of life.

Architecture Diagram



Dataset

This dataset shows crimes that happened in Los Angeles since 2020. It's based on paper reports, so there might be some mistakes. Some locations are marked as (0° , 0°) when data is missing. Addresses are given in approximate blocks for privacy. The accuracy depends on the database.

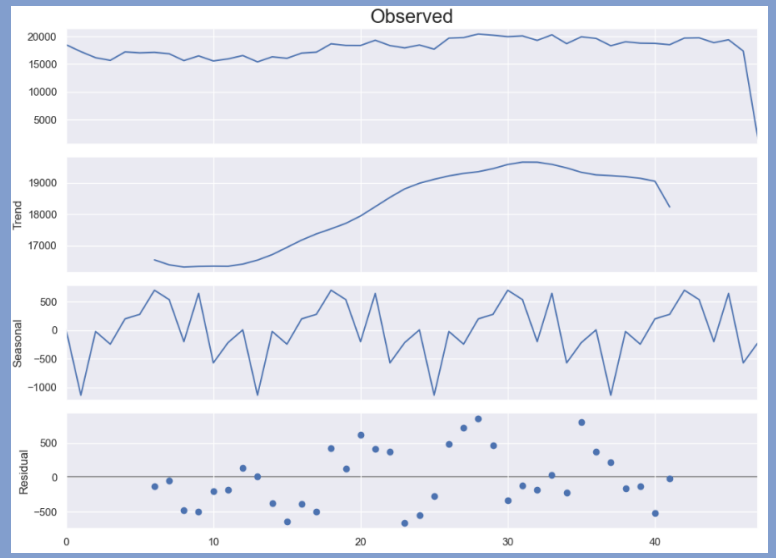
Last Updated: December 20, 2023
Data Provided By: Los Angeles Police Department

Columns in this Dataset

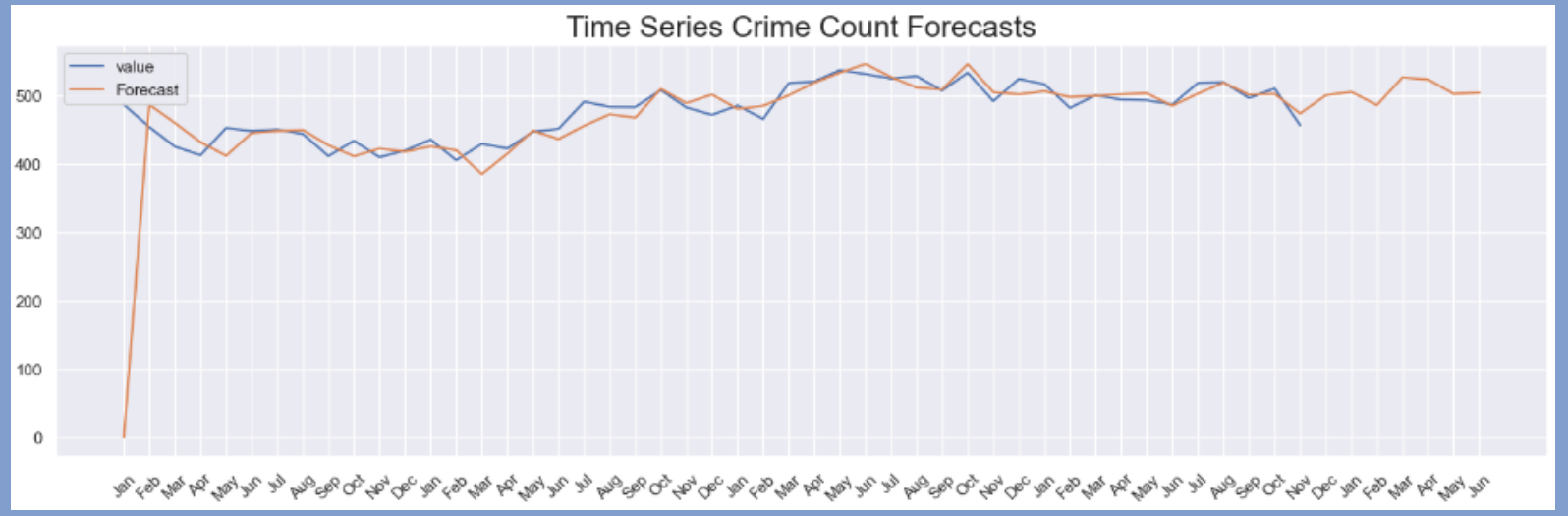
Column Name	Description	Type
DR_NO	Division of Records Number: Official file number made up of a...	Plain Text
Date Rptd	MM/DD/YYYY	Date & Time
DATE OCC	MM/DD/YYYY	Date & Time
TIME OCC	In 24 hour military time.	Plain Text
AREA	The LAPD has 21 Community Police Stations referred to as G...	Plain Text
AREA NAME	The 21 Geographic Areas or Patrol Divisions are also given a ...	Plain Text
Rpt Dist No	A four-digit code that represents a sub-area within a Geograp...	Plain Text
Part 1-2		Number
Crm Cd	Indicates the crime committed. (Same as Crime Code 1)	Plain Text
Crm Cd Desc	Defines the Crime Code provided.	Plain Text
Mocodes	Modus Operandi: Activities associated with the suspect in co...	Plain Text
Vict Age	Two character numeric	Plain Text
Vict Sex	F - Female M - Male X - Unknown	Plain Text
Vict Descent	Descent Code: A - Other Asian B - Black C - Chinese D - Camb...	Plain Text
Premis Cd	The type of structure, vehicle, or location where the crime too...	Number
Premis Desc	Defines the Premise Code provided.	Plain Text
Weapon Used Cd	The type of weapon used in the crime.	Plain Text
Weapon Desc	Defines the Weapon Used Code provided.	Plain Text
Status	Status of the case. (IC is the default)	Plain Text
elated Content		
Crm Cd 1	Indicates the crime committed. Crime Code 1 is the primary a...	Plain Text
Crm Cd 2	May contain a code for an additional crime, less serious than ...	Plain Text
Crm Cd 3	May contain a code for an additional crime, less serious than ...	Plain Text
Crm Cd 4	May contain a code for an additional crime, less serious than ...	Plain Text
LOCATION	Street address of crime incident rounded to the nearest hundr...	Plain Text
Cross Street	Cross Street of rounded Address	Plain Text

Link: https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data

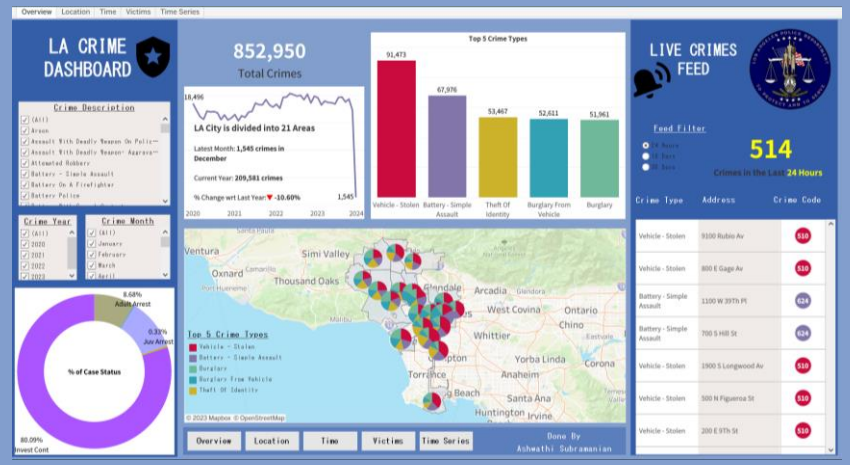
Time Series Analysis - SARIMAX

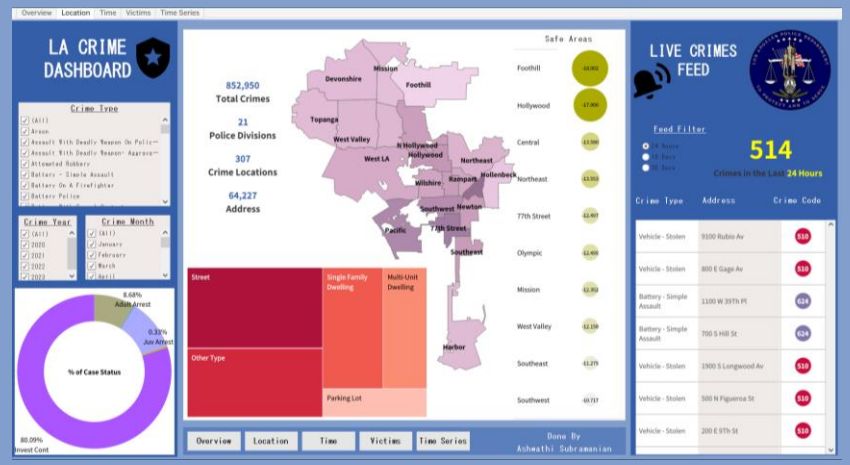


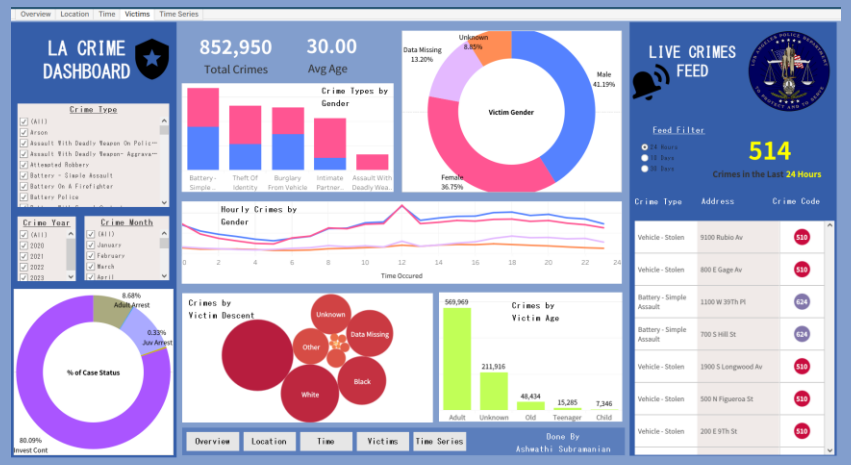
- Model Used: SARIMAX
- Order Parameters: (5,1,10)
- Seasonal Order Parameters: (6, 0, 0,12)
- RMSE: 95.09%



Demo







Link: https://us-west-2b.online.tableau.com/#/site/lacrimedataanalysis/home

Challenges Faced

- The project faced dataset challenges and chose LA crime data because there were no suitable APIs for Colorado.
- The Python versions for Spark and Airflow didn't match, so I had to downgrade both in the images, preventing the use of the latest Python version.
- In Tableau, despite drag-and-drop ease, view sizes varied between Desktop and Cloud. I used containers and reduced font sizes to ensure complete visibility on dashboards.
- The TSA Flint package, designed for time series in Spark, couldn't be used due to compatibility issues with other libraries such as NumPy and Pandas.

Future Work

1. Finding Safe Neighborhoods:
   - Help people discover safe neighborhoods near their desired address using recommendation systems.
   - Suggest areas with lower crime rates or higher safety levels for someone planning to move.

2. Safer Route Suggestions:
   - Offer alternative routes that avoid less safe areas, ensuring a safer journey.
   - Combine GPS and the model built to guide people away from places with higher safety concerns while traveling.

Thank You!!