

FINAL PROJECT PROPOSAL

DATA 603

BY: HAFSA CHAUDHRY

TABLE OF CONTEXT

* Introduction 1 slide

* Topic Domain Big Data Problem

1-2 slides

* Solution 1-3 slides

"Make sure to add if you have the data/where you got it from/ where you will get it from"

This PowerPoint is just a rough draft + outline for me.

1. INTRODUCTION

Topic: Los Angeles Crime Data

Overview: Reported Crimes in Los Angeles from 2020 to present

DATA SET FOR TOPIC

- 99.16 MB of data
- 4 files (3 PDFs and 1 CVS)
- 20 columns
- Data is categorized by codes (suspicious activity, traffic violation, etc) along with a tag number for each code describing which category the crime falls under (drunk, blind, mental illness, etc)
- Columns include information such as: date and time, age/sex/ethnicity, location, weapon, status of the case

Source:

- Kaggle: <https://www.kaggle.com/hemil26/crime-in-los-angeles>

2. TOPIC DOMAIN BIG DATA PROBLEM

Overview: crime rate prediction and analysis based on division and time

3. SOLUTION

Overview: Not really sure yet... Will throw brainstorm in next slide, please advise on scope and solutions

WHAT WE KNOW

- When crimes took place, where they took place, the victims descriptions, the type of crime and the case results

WHAT WE CAN DO WITH THAT INFO

- Visualize all reported crimes through a bar graph. Show which crimes are most common in GENERAL and/or visualize crimes per district

IDEAS

- We need to build some sort of analysis and prediction of the crimes...but what can we solve??
- Maybe have more LAPD presence at certain times in certain districts to monitor crimes and help in prevention?

HELP

On Kaggle, I can not see how many reports are listed or when they last updated. But they linked where they got the original dataset which is refreshed weekly and owned by the LAPD

They have listed 436k+ crimes in the dataset. How should I cut and slice this data for the scope of this course? And so my dingey slow no-storage computer can hold and handle

Thanks

9

FINISHED

Thank you!