



Pain Pills in Southern California

Presented by Group 1:

Alexander Castellanos, Vincent Cheung,
Steve Hwang, Eric Peralta, Armando Perez,
and Erick Robles



Overview

- Introduction to the Data
- Data Implementation
- Data Cleaning Process
- Questions About the Data
- Conclusion
- Q & A
- References



Introduction to the Data

- Collection of data from the DEA's database tracking many different pain pills sold in the US
- The Washington Post organized the transactions and separated them by state for ease of access for the public
- Transactions from 2006-2014





Introduction to the Data

- The unfiltered dataset for the entire US is roughly 80GB
- Narrowed down the dataset to southern California, reducing the size of the entire set to 4GB
- Focused on 11 different counties found throughout southern
- California

Data Set	Size (Total 4GB)
Arcos-ca-imperial-06025-itemized	28,420KB
Arcos-ca-inyo-06027-itemized	6,299KB
Arcos-ca-los-angeles-06037-itemized	1,658,659KB
Arcos-ca-mono-06051-itemized	2,222KB
Arcos-orange-06059-itemized	657,077KB
Arcos-ca-riverside-06065-itemized	477,904KB
Arcos-ca-san-bernardino-06071-itemized	425,997KB
Arcos-san-diego-06073-itemized	625,975KB
Arcos-ca-san-luis-obispo-06079-itemized	83,329
Arcos-ca-santa-barbar-06083-itemized	113,507KB
Arcos-ca-ventura-06111-itemized	198,779KB

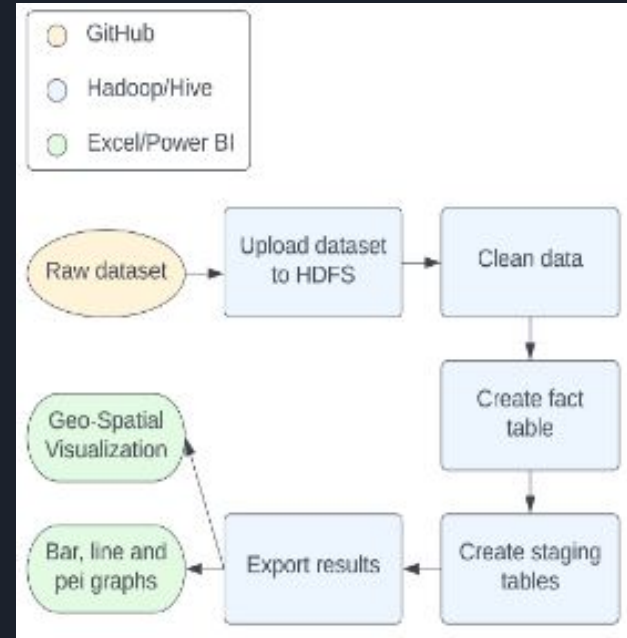
Introduction to the Data

- There were many columns with data in the original dataset
- Columns of interest:
 - Manufacturers
 - Distributors
 - Pharmacies
 - Products
 - Counts
 - Locations



Data Implementation

- Download raw dataset from Github
- Upload the dataset to HDFS
- Create a staging table for data cleaning
- After the data is cleaned, a fact table is created
- The table is then used for visualization





Data Implementation

- Hadoop Distributed File System specifications

Number of nodes	3
OCPUs	8
CPU speed	1995 MHz
Memory	58 GB
Storage	369 GB
HDFS Capacity	389 GB

Data Cleaning Process

- Raw data can often be out-of-date, denormalized, or poorly structured
- The data attributes are assessed to determine if they are numerical or categorical, while also assigning the level of importance to the variable
- The data is inspected for dirty entries
- A final table is created with clean, useful data



Data Cleaning Process

- Before data cleaning:
 - Dataset contained 42 columns
 - Transaction_data was a string data type
 - Drug_name and Zip columns had 507 bad entries
 - Manufacturer column had over 11,000 null entries



Data Cleaning Process

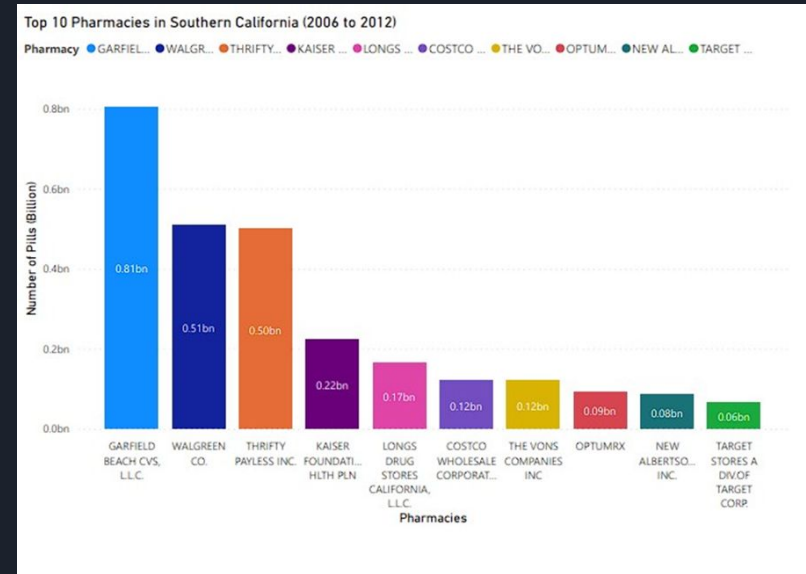
- After data cleaning:
 - Dataset contains 20 columns
 - Transaction_data is now a date data type
 - Drug_name and Zip columns have no bad entries
 - Manufacturer column have no null entries



Questions About The Data

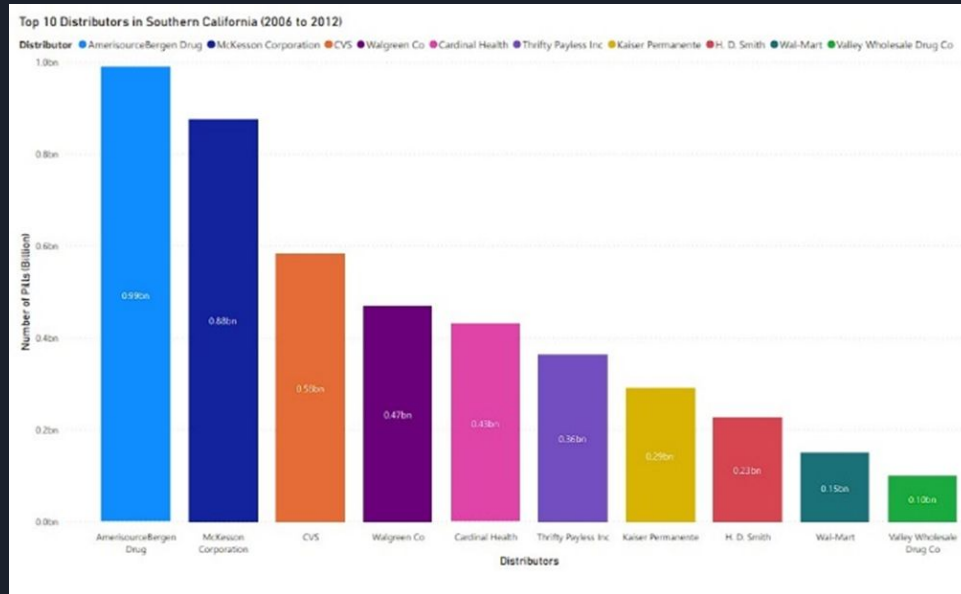
- **Who are the top 10 pharmacies?**

- Garfield Beach CVS
- Walgreens
- Thrifty Payless. INC
- Kaiser Foundation Health Plan
- Longs Drug Stores California
- Costco Wholesale
- The Vons Company
- OPTUMRX
- New Albertsons INC.
- Target Stores A DIV. of Target Corp.



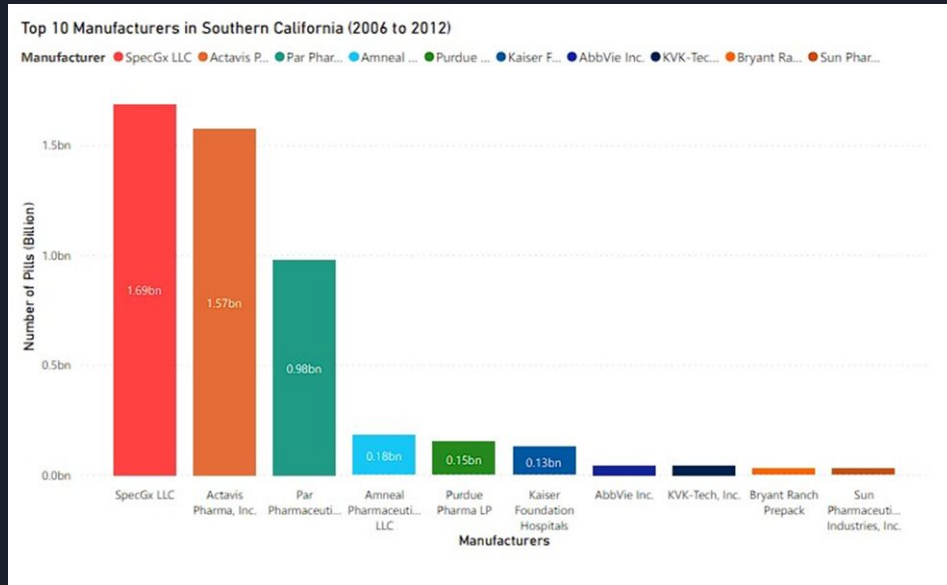
Questions About The Data

- **Who are the top 10 distributors?**
 - AmerisourceBergen Drug
 - McKesson Corporation
 - CVS
 - Walgreen Co.
 - Cardinal Health
 - Thrifty Payless. INC
 - Kaiser Permanente
 - H.D Smith
 - Walmart
 - Valley Wholesale Drug Co.



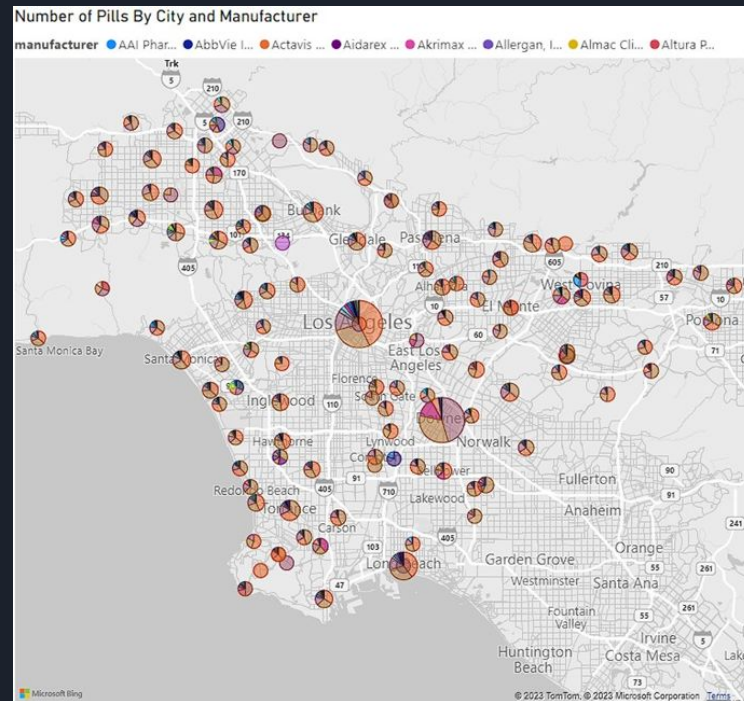
Questions About The Data

- **Who are the top 10 manufacturers?**
 - SpecGx
 - Actavis Pharma Inc.
 - Par Pharmaceuticals LLC
 - Amneal Pharmaceutical LLC
 - Purdue Pharma LP
 - Kaiser Foundations Hospital
 - Abbvie Inc
 - KVK Tech
 - Bryant Ranch Prepack
 - Sun Pharmaceutical Industries Inc.



Questions About The Data

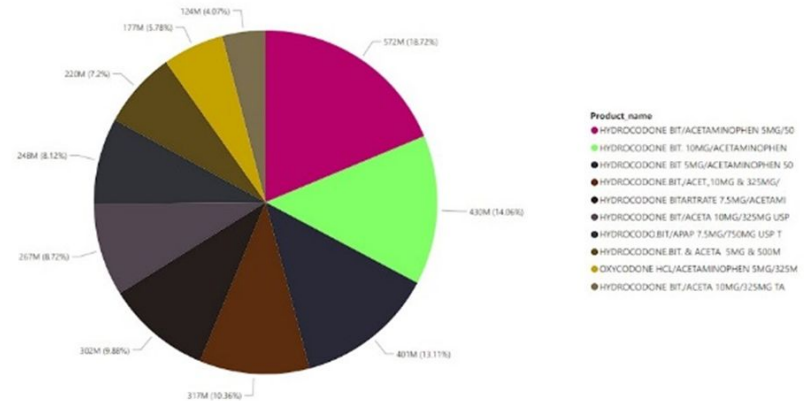
- Where are the majority of these painkillers manufactured in SoCal?
 - Los Angeles
 - Downey



Questions About The Data

- What is the most popular pain killer?
 - Hydrocodone Bit/Acet. 5mg/50
 - Accounting for 18.72% of the pain pill market.

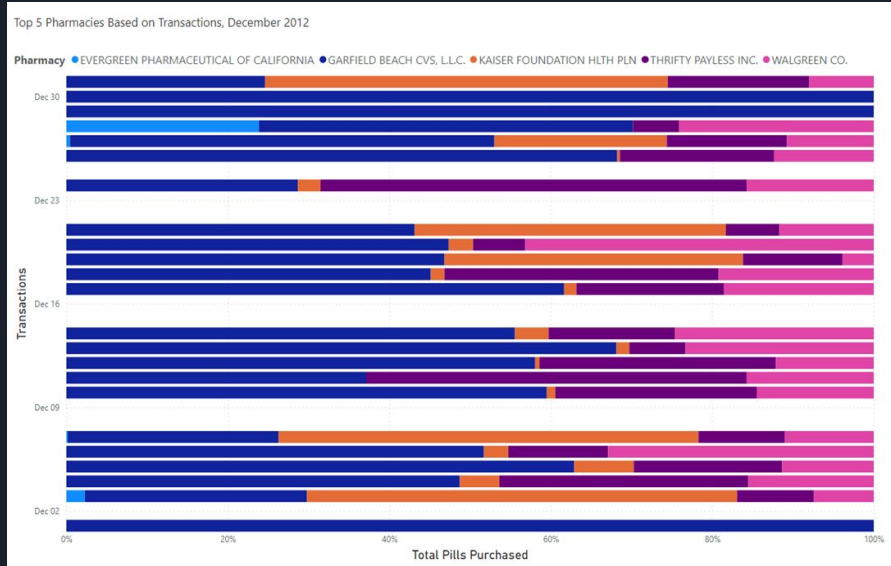
Top 10 Products in Southern California (2006 - 2012)



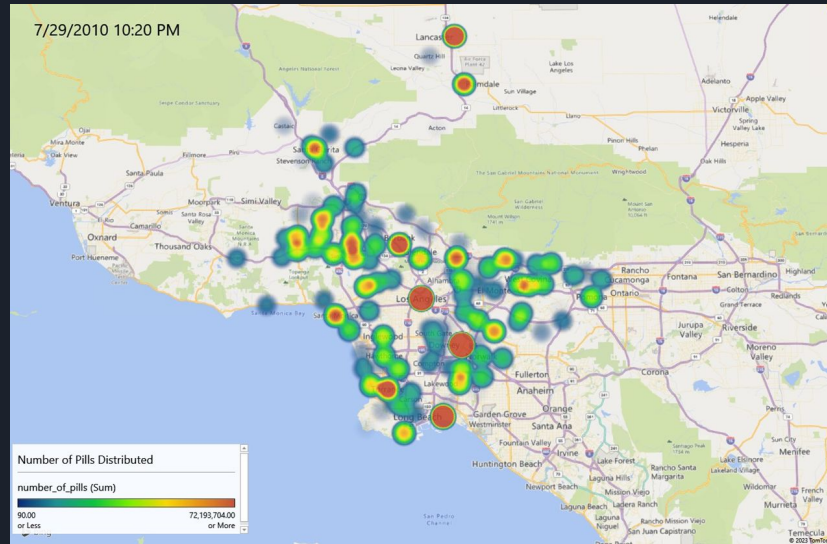
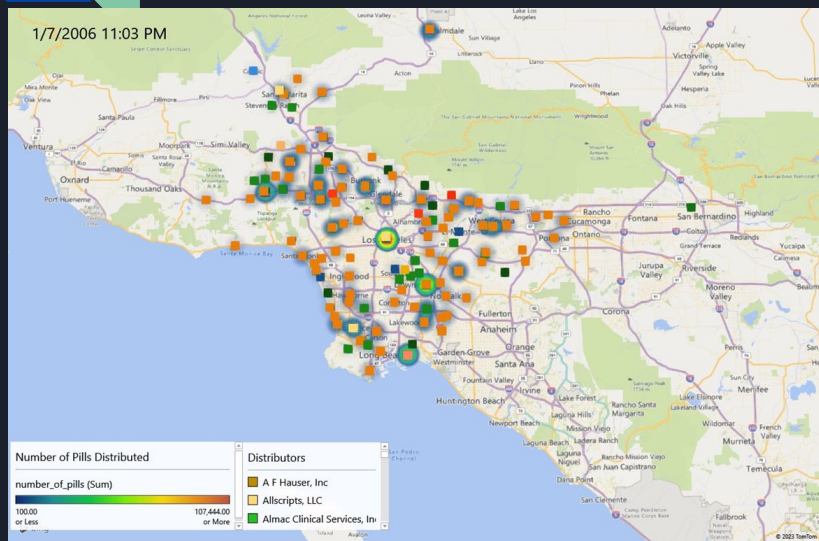
Questions About The Data

- **Who are the top 5 pharmacies based on transactions?**

- Walgreens
- Thrifty Payless
- Kaiser Permanente
- CVS
- Evergreen Pharmaceutical



Questions About The Data



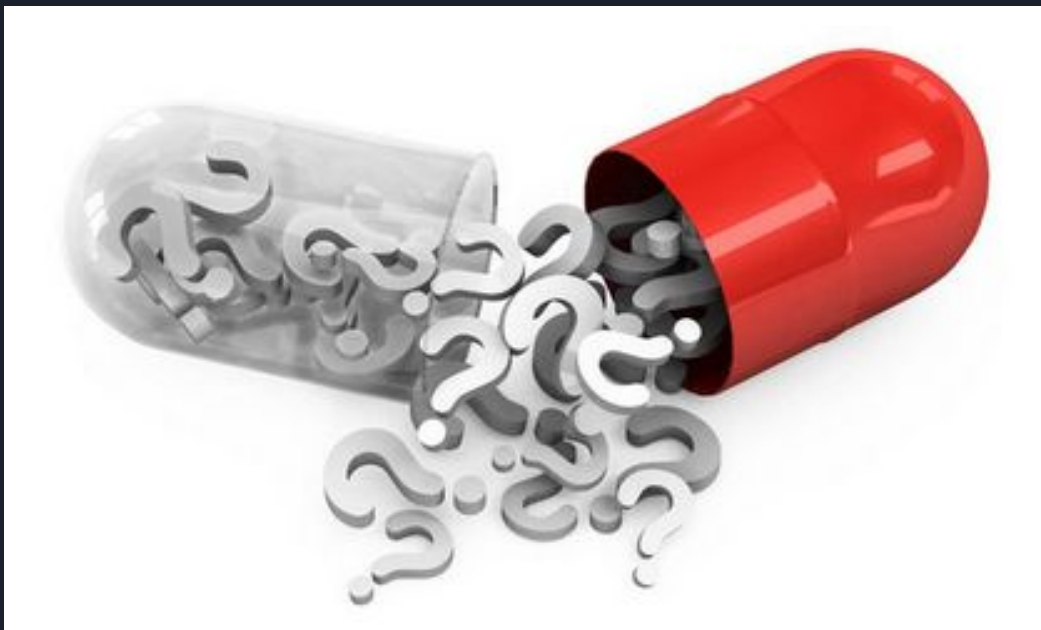
- Where in SoCal are these products most popular?
 - Downtown Los Angeles
 - Central Los Angeles

Conclusion

- Based on evidence above, we can conclude that:
 - Opioid distribution is concentrated in densely populated areas of LA County such as Downtown Los Angeles and Central Los Angeles
 - Pharmacies increased purchases after 2010
 - Manufacturing is concentrated in DTLA and Central LA
 - The majority of opioids produced are hydrocodone and oxycodone



Q & A





References

[1] Pain Pill Database (July 2019) DEA - ARCOS.

Retrieved from t.ly/O_SC

[2] California Overdose Surveillance Dashboard (February 2023) – Skylab

Retrieved from:

<https://skylab.cdph.ca.gov/ODdash/?tab=Home>

[3] OPIOID_DATA_ARCOS_DEA_IN.GDB: Opioid Pill Distribution Information for Indiana from 2006 to 2012, from the U.S. DEA's Automation of Reports and Consolidated Orders System (ARCOS), (DEA, The Washington Post, File Geodatabase) (August 2018) – Indiana Geological and Water Survey

Retrieved from t.ly/Cenj

[4] Github: <https://github.com/vcheung621/cis4560>