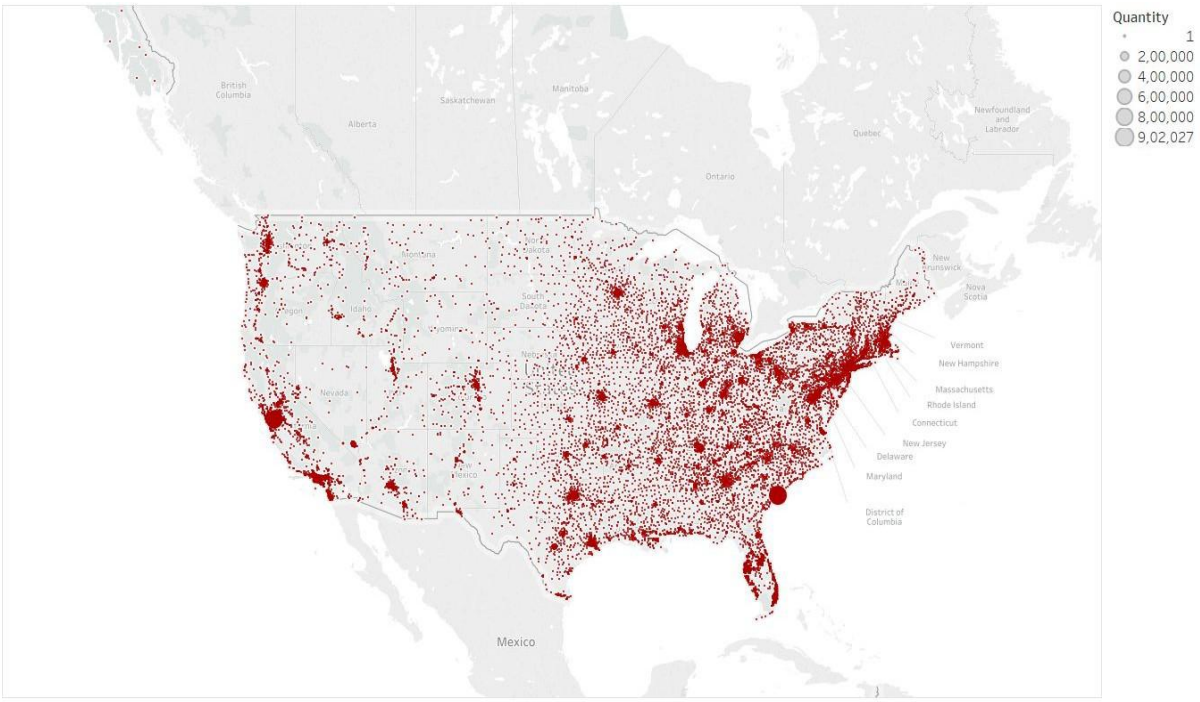
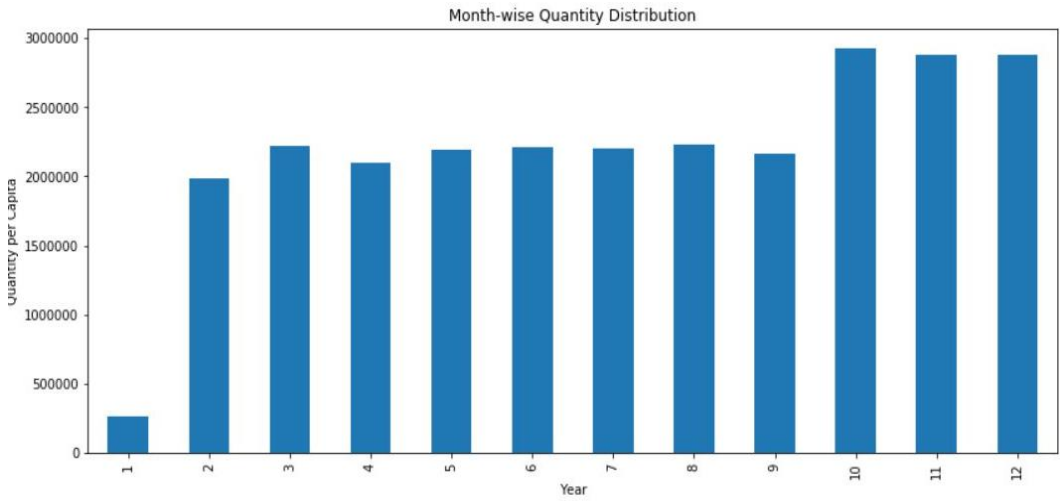
Understanding the Opioid Epidemic in the United States

The rise of deaths due to opioid overdose has become an epidemic in the United States. Drug overdoses are the leading cause of death for Americans, who are now more likely to die from a drug overdose than from car accidents or firearms. The US has the dubious distinction of having the highest percentage of drug-related deaths in the world. While opioid abuse is a nationwide problem, there are specific areas that are being hit harder by this epidemic. The below map shows zip codes and their related opioid sales.



Map based on Longitude (generated) and Latitude (generated). Size shows sum of Quantity. Details are shown for Buyer Zip.

It is vital to address the issue and understand the in-depth reasons for abuse of these drugs. Our initial hypothesis was that age may be a significant factor contributing to the overall drug abuse. We tested this hypothesis with some exploratory data analysis of the available data along with some implications from external sources and found it to be false. When we saw this, we set out to determine what we could learn by analyzing the distribution data. To accomplish this, we analyzed the trend, patterns, and seasonality in the Opioid Pill distribution on daily, monthly, and yearly basis.



The monthly plot shows 3 distinct groups of distribution as in the above figure. The Distribution in the months of October, November, and December is about 50 percent higher than the mean distribution from February to September. On the other hand, the distribution is drastically low in case of January which might be attributed to the consequences of slow business in the holiday season. We can infer from the increasing distribution of opioids in the winter season that the risk of fatal opioid overdose increases. The same has been validated by the research done by Professor Marshall at Brown University, wherein he analyzed the deaths associated with opioid overdose in Rhode Island and Connecticut and concluded that the cold snaps raised such a risk by about 25 percent.

America's largest drug companies and pharmacies have saturated the country with 76 billion oxycodone and hydrocodone pain pills from 2006 through 2012. The information comes from Automated Reports and Consolidated Ordering System (ARCOS) database which is maintained by the Drug Enforcement Administration (DEA). The ARCOS tracks the path of every single pill sold in the states - from manufacturers and distributors to pharmacies in each and every town and city.

Upon analysis of this data, we found that the top 5 pill purchasers to be: CVS, Walgreens, Walmart, Rite Aid, and Safeway as seen in the pie chart to the left. This chart helps us understand some of the key players in this epidemic

at a consumer level which has resulted in nearly 100,000 deaths from 2006 through 2012. We also looked at their opioid sales per retail location and compared that with the average per consumer retail location as seen in the bar chart graph to the left.

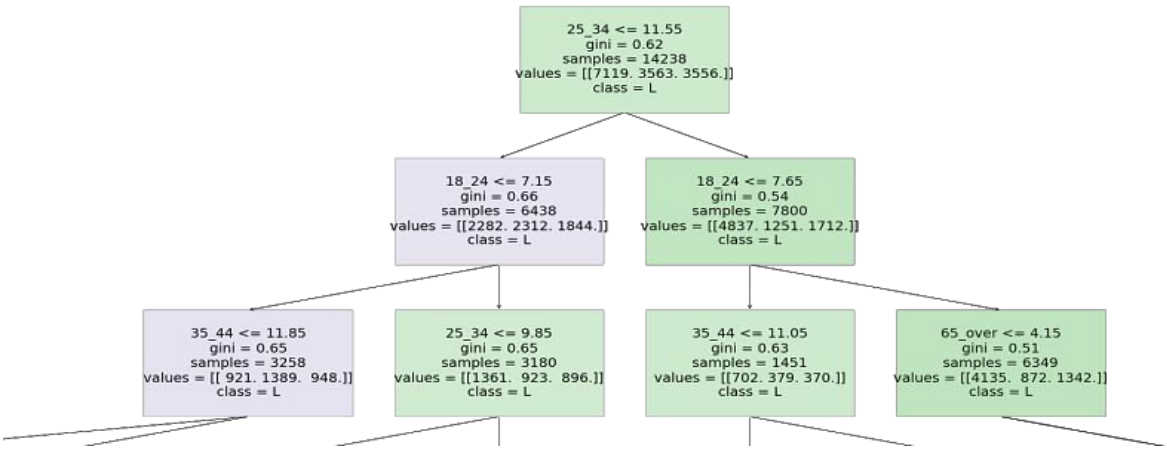
The raw data is based on shipments of oxycodone and hydrocodone pills to chain pharmacies, retail pharmacies and practitioners. However, the original data included 10 other opioids which were removed as they were shipped in much lower quantities and were diverted at far lower rates over the seven years.

We can see that there is a particular group of pharmacies that dominate this market and necessary actions should be taken by statutory authorities in order to curb the sale of pills. The sale of pills is concentrated in most of the north-eastern cities and all major metropolitan cities. However, it cannot be determined from this dataset whether the consumption of pills is also centered around these locations or these are just intermediary transfer points.

Mining pattern in certain age groups

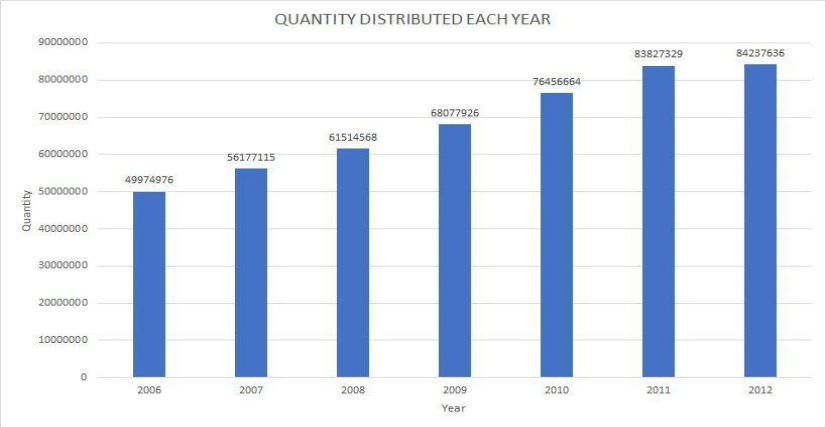
The idea behind mining patterns in different age groups along with their consumption was to verify if there is any particular group that tends to consume more number of pills on an average. For example, if a zip code has more population in the age group of over 65 years, then it is likely that they might need more medication due to aging and lower body tolerance. Such a group is more prone to be affected by drug overdose. Along with this group, if you consider youth in the age group of say 16-30 who are empirically more prone to overdose, might be a reason of over selling of these opioid pills in that particular zip area. If we are able to find any particular age group is affected more by this crisis, we can leverage this information to identify some zip codes and then the law enforcement could be alerted about the same. They can take the necessary precautionary measures such as prescription of low dosage unit, track illegal drug distribution, etc. We modelled a decision tree to test the hypothesis, 'Whether a particular combination of age groups affect the chances of drug overdose?' because a decision tree would be a good model to analyze and visualize any hidden combinations of variables that might be leading to a particular target.

We decided to use the 2010 US census data and filtered the original dataset by all the buyers aggregated over their zip codes. While preprocessing, we dropped the counties or zip codes which were missing from either of the datasets. Moreover, some of the outliers such as zip codes 29405, 94550, 90242, and 37129 were removed as well. The quartile range of the pills sold were found and any value below 25th percentile was labelled as low, any value above 75th percentile as high, and the interquartile range was as medium. From the Decision Tree Model, we found out that there is no discernable pattern between the pill distribution and any particular age groups.

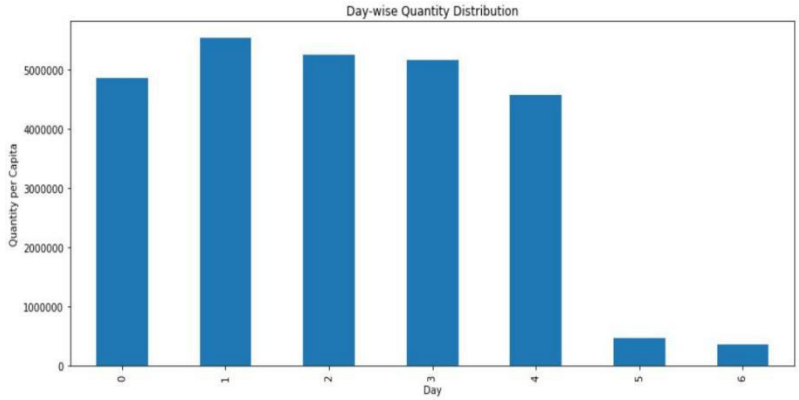


A Time Series Analysis approach

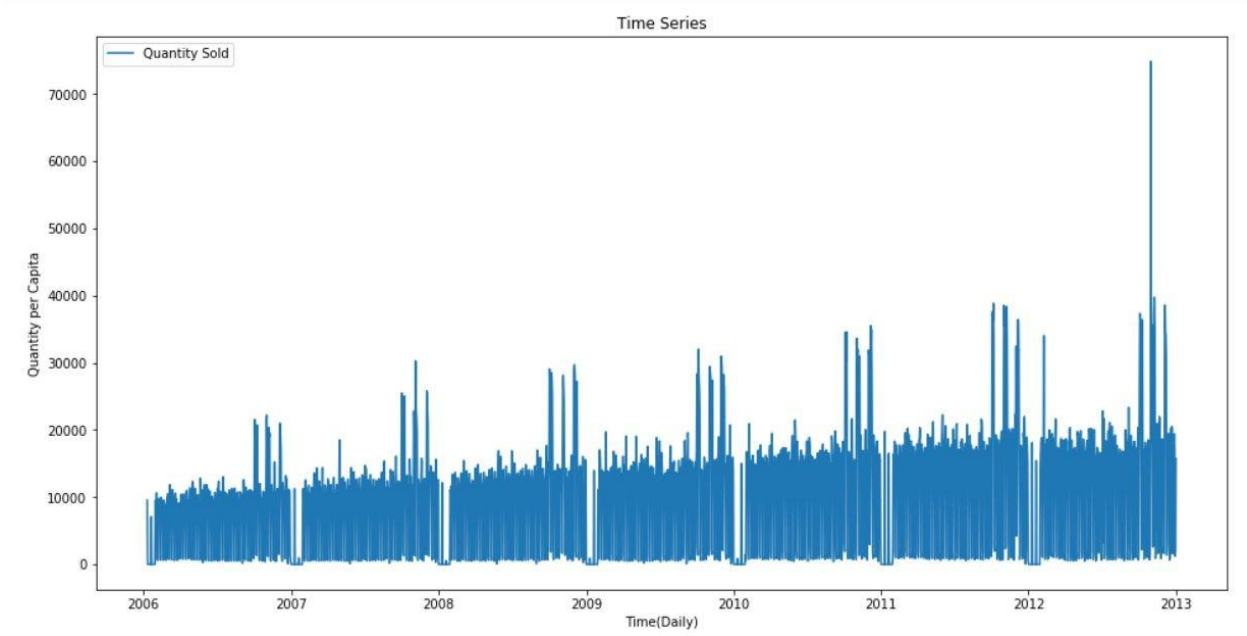
Using the 'TRANSACTION_DATE' information available in the ARCOS dataset, we studied the buying patterns and trends on daily, monthly, and yearly basis. For this analysis, we used only twenty million rows of data because of computing limitations of our machine. Following trends were observed during the exploratory data analysis of quantity of opioid distribution over different time stamps.



It is clear from the figure that the opioid distribution has consistently increased ever since 2006 indicating that a growing prescription trend among the Doctors and the promotion by the manufacturers.



The day-wise trend shows approximately 80% lower figures for weekends which seem to be due to less or no business hours for the distributors on weekends.



An interesting pattern was observed while observing the time series plot for the Total Quantity sold on a daily basis for 7 years from 2006 through 2012. There is evidence of strong seasonality in this case as we can see three distinct spikes every year in the last quarter. It is an indication that the sales of these opioids are being further pushed or promoted by the manufacturers and distributors to meet their financial goals for the fiscal year. Such hypothesis is an alarming sign from the manufacturer as it conveys their focus on profitability than on the public health and safety.

References

1. Cold weather increases the risk of fatal opioid overdoses, Brown University, July 2019.
2. Washington Post
3. The Automated Reports and Consolidated Ordering System (ARCOS) by the Drug Enforcement Administration (DEA)
4. Kaggle.com