**INFORMATION SEEKING**

**Dataset 1 – Traffic Violations**

**APA-formatted data citation**

Montgomery County Police. (2016). *Traffic Violations* [Data file]. Retrieved from https://data.montgomerycountymd.gov/Public-Safety/Traffic-Violations/4mse-ku6q

**Terms of use**

This dataset is intended for public access and use. Use of data from the Montgomery County website is subject to the terms of use provided in the following URL:

https://data.montgomerycountymd.gov/terms-of-use

The chosen dataset contains traffic violation information from all electronic traffic violations issued in Montgomery County. With attributes such as date and time of the violations issued, location, vehicle type, vehicle model, violation charge code, driver race and gender, it is possible to analyze various violation patterns by month, vehicle, race and gender. Furthermore, the data is well structured and provided in a simple and easily understandable format. The dataset is regularly updated with the last update made on September 15, 2016, which makes our statistical analysis relevant to the current time period. With recent allegations of discrimination by police officers doing the rounds, it may also be possible to verify the same using the data provided, by checking for possible patterns of bias by police officers. The possibilities of exploring and quantifying a variety of relationships and theories by utilizing simple, yet effective data is what appealed to me the most.

**Potential data users and decision makers**

* The violation frequency patterns observed in the data can aid the Montgomery County Police sub agencies in decision making.
* Accident and Damage assessment information at the different county locations can aid Insurance companies frame relevant policy packages for residents.
* Drivers can be made aware of possible high-risk locations in the county through mobile apps designed using the violation patterns observed in the data.

**Three questions this data might help to answer**

* At what time periods of the day are county policy officers of every sub-agency most likely to be busy due to increased violation frequencies?
* Do vehicles of a specific model or color contribute more towards a particular violation type?
* Are specific violations in a particular location caused more by drivers of a particular race? Can a bias in the handling of these violations be detected?

**Dataset 2 – Consumer Complaint Database**

**APA-formatted data citation**

Bureau of Consumer Financial Protection. (2015). *Consumer Complaint Database* [Data File]. Retrieved from https://catalog.data.gov/dataset/consumer-complaint-database

**License**

No license is required. The following notice is available in the website of the Bureau of Consumer Financial Protection, the link for which is provided below.

‘Website content: Information created by the Consumer Financial Protection Bureau is in the public domain and you may reproduce, publish, or otherwise use it without the Bureau’s permission. Please consider appropriate citation to the Bureau as the source.’

Link: http://www.consumerfinance.gov/privacy/digital-privacy-policy/

In today’s digital age, there is a growing importance given to the handling of consumer complaints by businesses. The dataset selected contains a number of comprehensive parameters to judge the effectiveness with which a company handled customer grievances. Analyzing patterns in this data can enable companies to design better customer support systems. Both quantitative and multiple choice categorical data are available for analysis. Data is provided for a number of product and sub-product categories as well as complaint channels thereby increasing the range of our analysis.

**Potential data users and decision makers**

* The analysis of data in this dataset can be used by companies to plug in loop-holes in their customer service process.
* It can be used to identify the most effective channel in redressing customer complaints.
* Customers can review the complaint frequency patterns for different products in order to make an informed decision before purchasing a particular product.

**Three questions this data might help to answer**

* Which company provides the best customer complaint redressal service in every product category?
* Which channel proves most effective in dealing with customer complaints every year?
* Maximum number of complaints were lodged against which product category when ordered by states?

**Dataset 3 – Amazon Fine Food Reviews**

**APA-formatted data citation**

Stanford Network Analysis Platform. *Amazon Fine Food Reviews* [Data file]. Retrieved from https://www.kaggle.com/snap/amazon-fine-food-reviews

**License Details**

This dataset was released under a [CC0: Public Domain License](https://creativecommons.org/publicdomain/zero/1.0/). It gives the user the rights to copy, modify and distribute the work, even for commercial purposes without seeking permissions.

Additional Information

* In no way are the patent or trademark rights of any person affected by CC0, nor are the rights that other persons may have in the work or in how the work is used, such as [publicity or privacy](http://wiki.creativecommons.org/Frequently_Asked_Questions#When_are_publicity_rights_relevant.3F) rights.
* Unless expressly stated otherwise, the person who associated a work with this deed makes no warranties about the work, and disclaims liability for all uses of the work, to the fullest extent permitted by applicable law.
* When using or citing the work, one should not imply [endorsement](https://creativecommons.org/publicdomain/zero/1.0/) by the author or the affirmer.

Links

* https://www.kaggle.com/snap/amazon-fine-food-reviews
* https://creativecommons.org/publicdomain/zero/1.0/

A glut of information about products and services in the form of reviews are available on websites, enabling consumers to make informed decisions. In this dataset, a score rating is assigned to every food product. A unique user-id is assigned to every user, which enables us to track the reviews posted by the user for multiple products. Thus a relationship between user activity and helpfulness of the product review can be established by analyzing the data. Patterns observed in this dataset can test the degree to which food review sites are helpful.

**Potential data users and decision makers**

* Analysis performed on the data can help customers select/ignore a particular product based on its review score rating.
* Companies can analyze their product performance over a period of time
* Amazon can develop algorithms to reduce fake reviews based on data analysis.

**Three questions this data might help to answer**

* Did the ratings for a particular product improve or depreciate over a period of time?
* How many individuals who reviewed the products posted a fake review?
* Which food product has been reviewed the most?