**Name:** Don Price

**Class:** DSC 540

**Assignment:** Final Project

**Understanding Flood Risks in Texas**

Flood losses are a recurring risk to homeowners and businesses, especially those located near coastal areas. To understand these risks, I constructed a database containing information about the risks and exposures within Texas. The database contains:

* National Flood Insurance Program (NFIP) Claims – a csv file from NFIP
* NFIP Policy Data – obtained from an API
* Housing totals by county – risk data scraped from a government website.

This database allows users to understand the magnitude of insured risks in a geographic area compared to the total housing population. Presumably, a place with higher flood risks will have a greater proportion of insureds. Using claims data, we can assess the level of risk in an area and measure the significance of losses. The graphical analysis in my notebook analyzes risk and losses in TX over the past five years. This analysis can be useful to homeowners needing to mitigate risks of flood and government authorities who need to set a fair price for flood insurance.

**Project Learnings**

Consolidating the data from differing sources identified several counties at high risk of flooding, based on historical losses. In these areas, flood insurance is a commonly purchased product with 20-40% of homes purchasing flood insurance in risky areas. In areas with high historical flood losses, all homeowners should carry flood insurance. This is something government authorities should seek to correct through awareness campaigns or mandates. Profitability of the program is strong for the 4 years where we have both premium and loss data, although it would be desirable to measure this over longer periods of time for a proper assessment.

**Ethical Implications of Project**

The NFIP database is the primary dataset. This data set contains all policies and claims issued by the NFIP. Information about the policy, covered risks, and location of risk is available. No personally identifiable information is present, so no data sharing or exposure risks are present.

Use of the data and analysis allows government agencies to be informed on the risks of flood, purchase patterns of homeowners, and profitability of the program. Homeowners can use this information to understand the risks of flood and to be informed when deciding whether to purchase flood insurance. Usage fits within the guidelines of the NFIP and would be in accordance with the privacy regulations.

**Personal Learnings**

The project allowed me to explore new methods for acquiring data. My historical experience is with relational databases that are highly structured. This project allowed me to deeply explore web scraping, API usage, and gathering a flat file from the web. Through processing this information, I gained strong experience in using Python to organize information. This skill will be useful in the real world as we build data pipelines for our data science models. Building a SQL database was a lot of fun and is much less intimidating than it originally seemed. I have some experience with SQL already, but I’ve not used Python and SQL interchangeably, which is needed for constructing a data science pipeline.

Through this project, I have gained significant confidence in preparing data for modeling and other uses. I have also gained a deeper understanding of data sources, which has broadened the ideas that I may have for professional projects moving forward. I no longer feel constrained to a relational database.