



# SafeHouse NYC

A Real Estate Marketplace  
with Crime Insights

Managing Data Term Project By:

Aadhar Kaul (ak4937)

Abdulaziz Alshammari (afa2150)

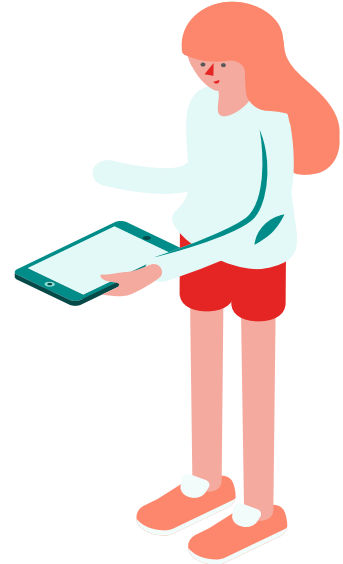
Saad Hazari (sh4377)

Saloni Dingore (srd2180)

Srushti Divyesh Sanghavi (ss6594)

# Customer value Proposition

- Provides real estate listings and neighborhood crime statistics, helps make informed decisions
- Increasing transparency and awareness around crime statistics and home values
- Helping users feel more comfortable and confident in their choice of living



# Literature Review

---

- Location, size of property, state, perks, economy, housing regulations, and demographic influence customers' decisions when it comes to real estate.
- Crime negatively impacts one's decision of choosing a property, with different types of crimes having varying impacts on customer sentiments
- Decreased crime rates have a positive effect on one's choice of choosing a property, as shown by studies.
- Other variables, such as socioeconomic factors and house location affects the dwelling preference of customers, with the level of perceived safety and desirability decreasing as crime rates increase.



# Data Sources

---



For the purpose of this project, we are using Zillow real estate data and NYPD Crime complaint data. Due to API access limitations, we are using the datasets available on Kaggle.

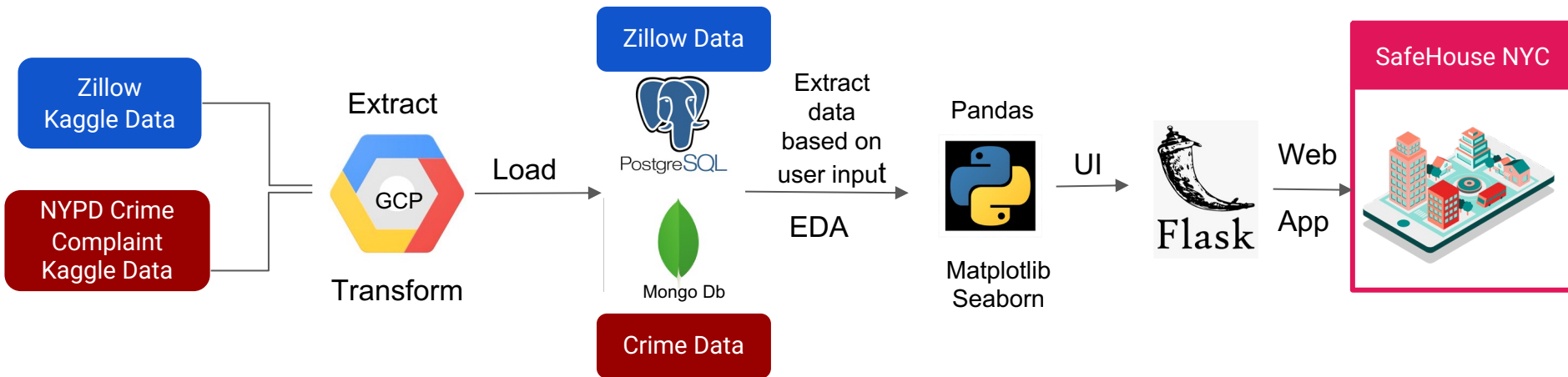
The datasets are available in the form of a csv file.

The data sources are as below:

1. [Zillow New York Housing Data](#)
2. [NYPD Crime Complaint Data](#)

We plan to integrate Zillow listing data with different types of complaint data such as shootings based on latitude and longitude, which are available in both datasets.

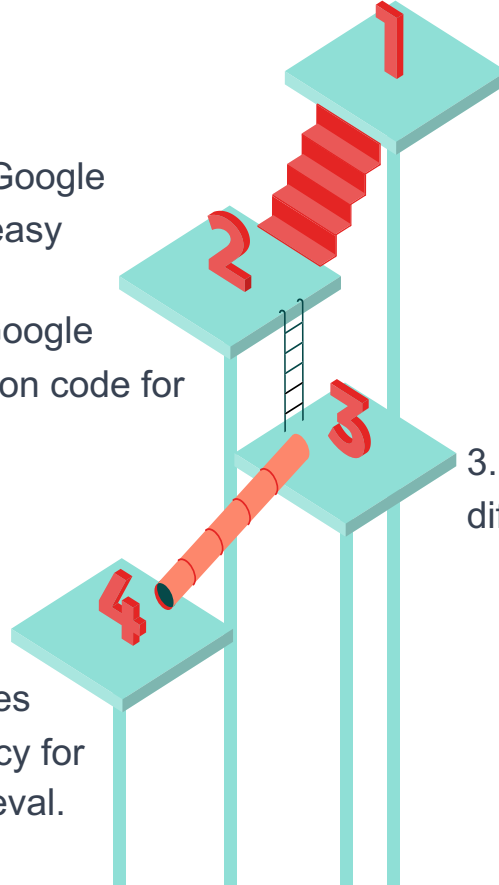
# Flowchart



# ETL

2. Store the extracted data into a Google Cloud Storage (GCS) bucket for easy access and management.  
Used Vertex AI environment on Google Cloud Platform (GCP) to run Python code for cleaning and transformation.

4. Create tables in the databases based on functional dependency for efficient data storage and retrieval.



1. Extract data from two sources:  
NYPD crime dataset from Kaggle  
and Zillow API crime dataset.

3. Load the transformed data into two different databases:

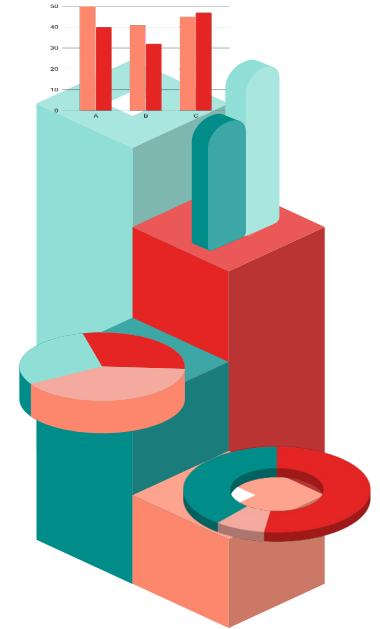
- Store the cleaned Zillow data into PostgreSQL.
- Store the crime data with zipcode information into MongoDB.

# EDA

---

After performing the Extract-Transform-Load (ETL) process on the NYPD crime dataset and Zillow API crime dataset, we conducted exploratory data analysis on the combined dataset to gain insights and identify trends. Below are the type of graphs we have incorporated:-

1. Borough-wise Distribution
2. Offense Type Distribution
3. Suspect Demographics:
4. Premise Type Distribution:



# Flask

- Our app SafeHouse NYC displays a list of available Zillow listings for the input zipcode.
- The user can input a zipcode and get information on crime statistics and Zillow listings for that particular zipcode.
- We are using flask framework and html to build our web application .





# Scalability



## 1. Infrastructure and Architecture:

- Cloud computing and infrastructure scaling
- Distributed systems architecture
- Load balancing and traffic management
- Database sharding and partitioning

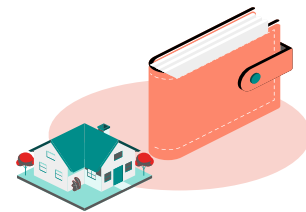
## 2. Performance and Optimization:

- Caching and database optimization
- Application performance management and optimization
- API design and management for scalability and flexibility
- Monitoring and analytics for performance and scalability

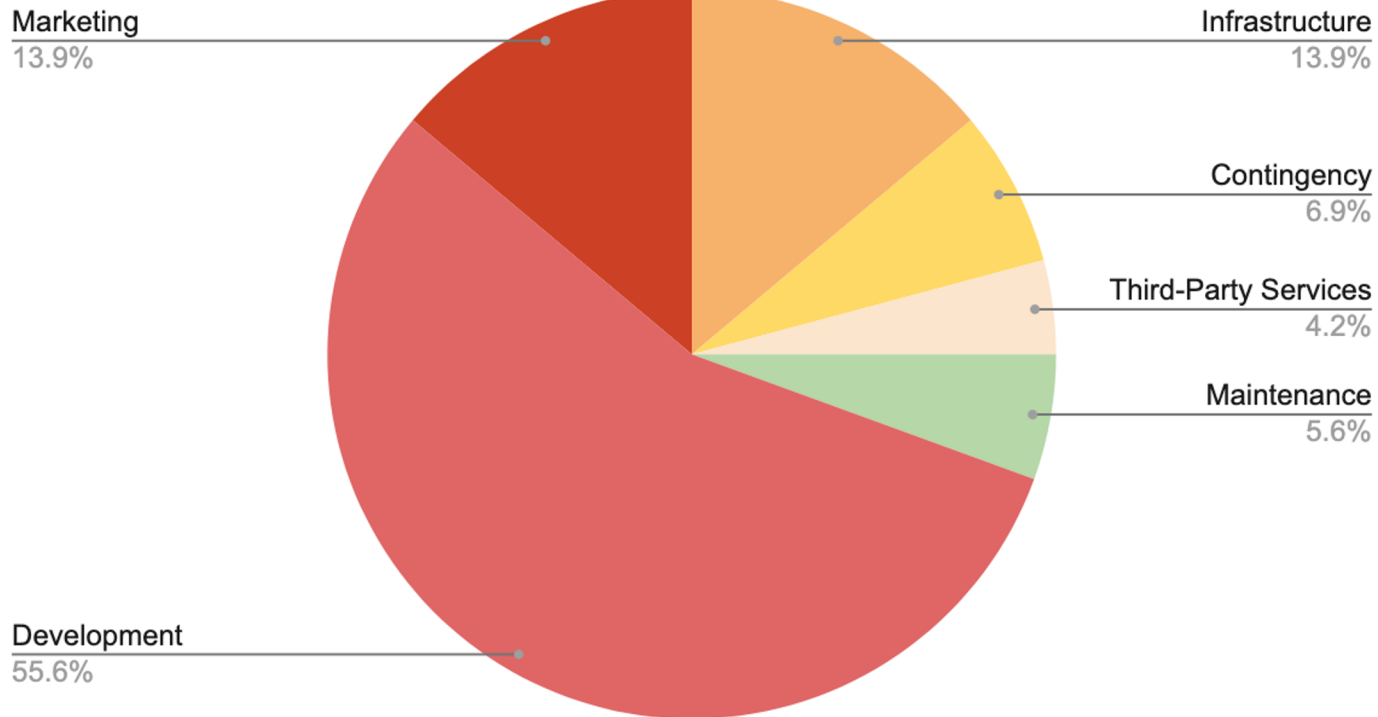
## 3. Advanced Techniques:

- Big data processing and analysis
- Machine learning and artificial intelligence for predictive scaling
- Failover and disaster recovery strategies
- Horizontal and vertical scaling strategies

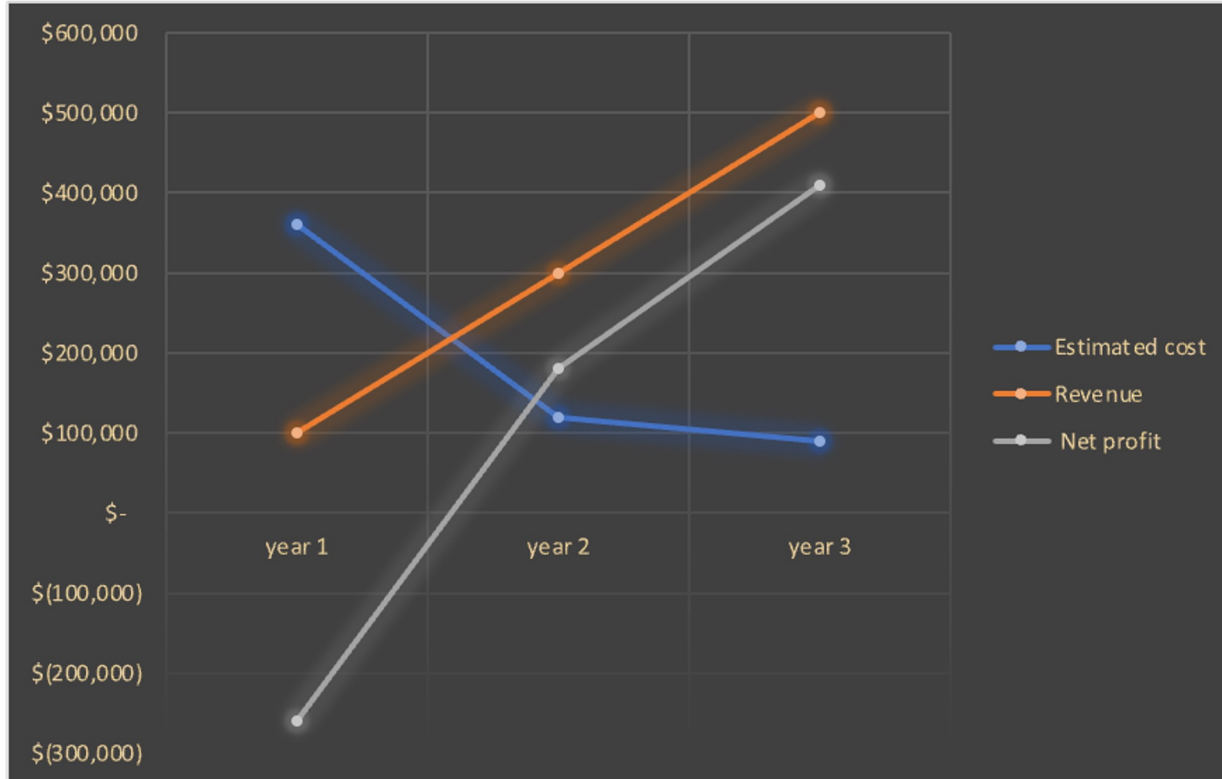
# Cost of the project



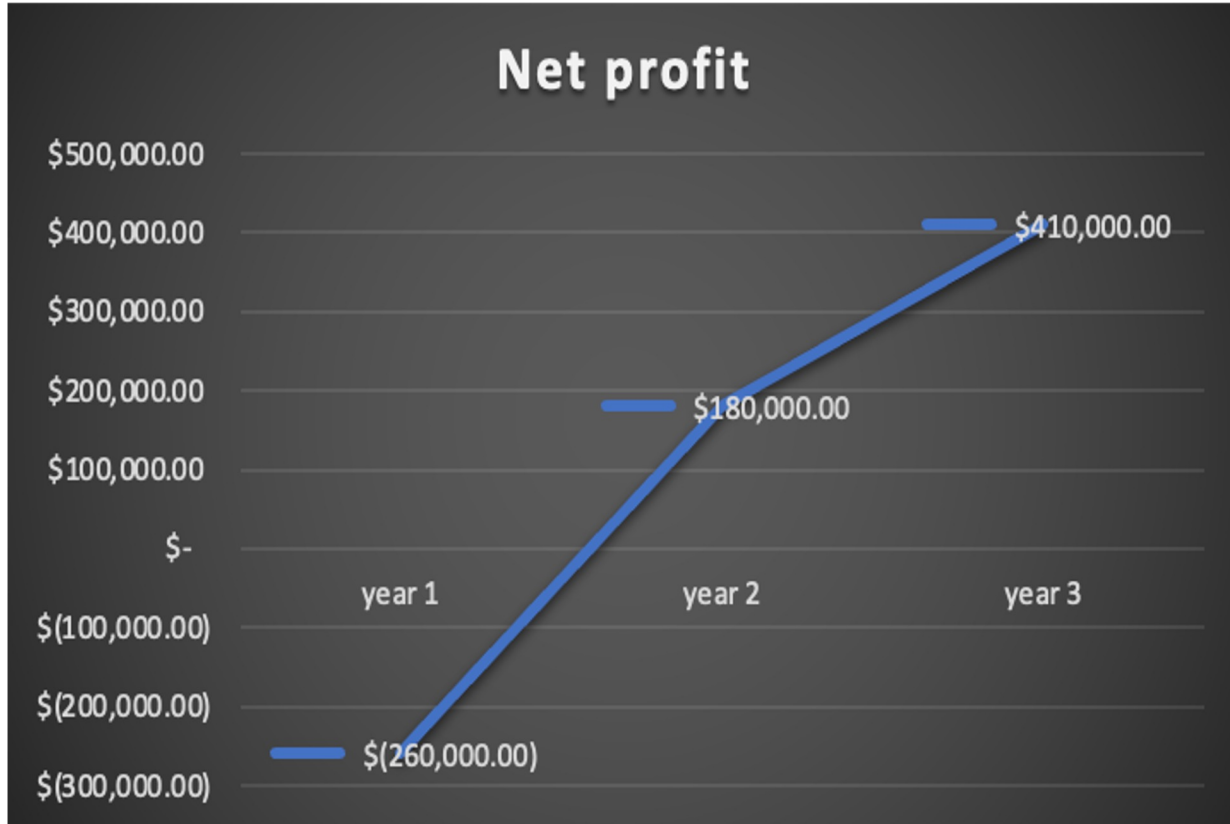
## Estimated Cost



# Revenue model



# Net Profit



# Limitations

---

- Availability of Data: Access and up to date available data on crime statistics and real estate data could be limited or incomplete, which could affect the usefulness of the app.
- Privacy Concerns: The app may face privacy concerns as it may collect personal data, location information, and browsing history of the users.
- Technical Challenges: The app may face technical challenges, such as compatibility issues with various devices, slow data processing, or data analysis errors, which may hinder the overall user experience.



THANK YOU

