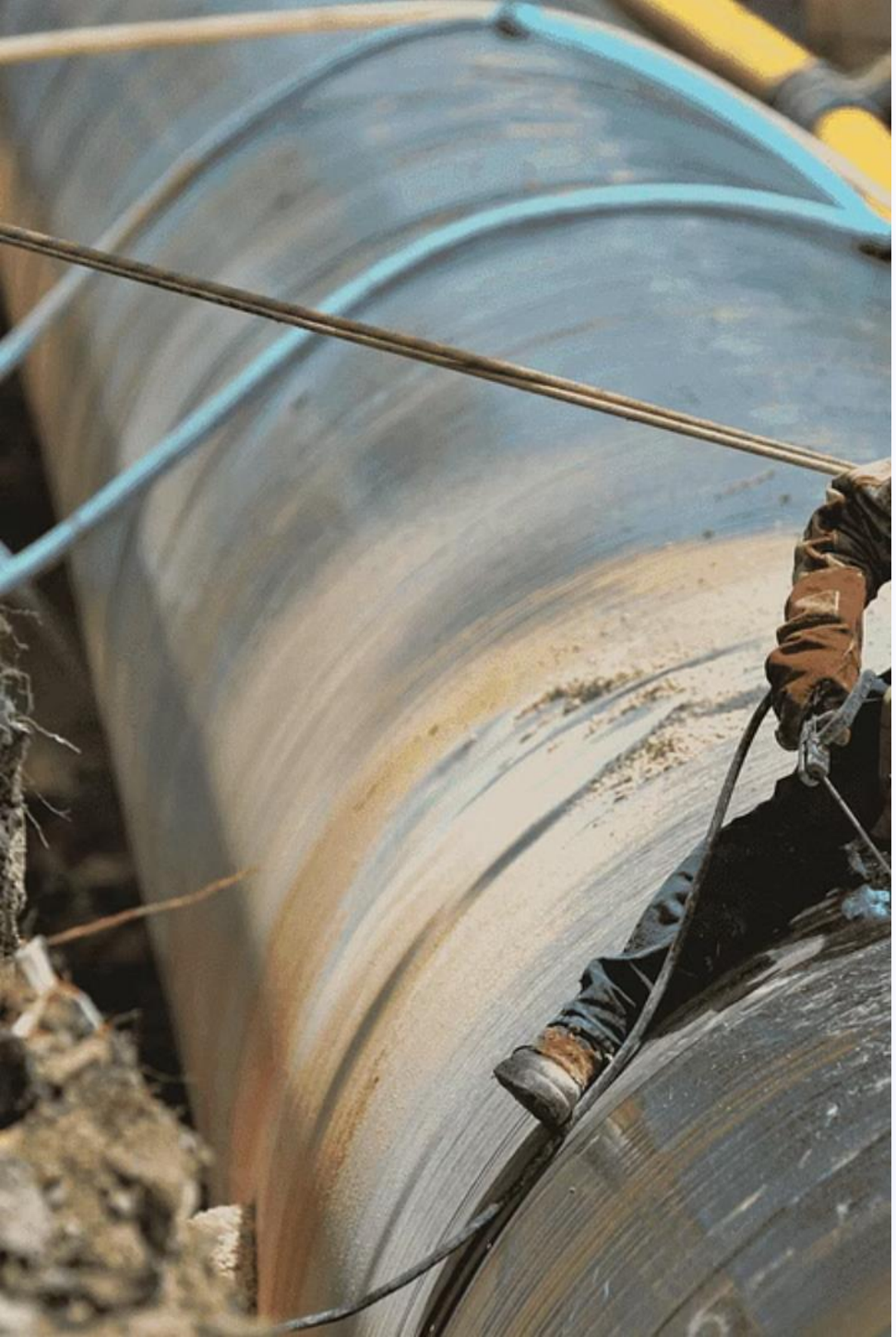


# Report - Hackathon

Strategic deployment of straight-line segments for gas gas pipeline construction, aiming to efficiently serve n serve n houses while minimizing environmental impact impact and construction costs.

By Team - **Underdogs**

# Team Members -

Daksh Jain ( Team Leader )

Siddharth Gupta

Aamir Imam

Faizan E Rasool

# Challenges encountered:

1. **Cluster Representation:** Ensuring accurate representation of houses through latitude and longitude coordinates in the clustering process.

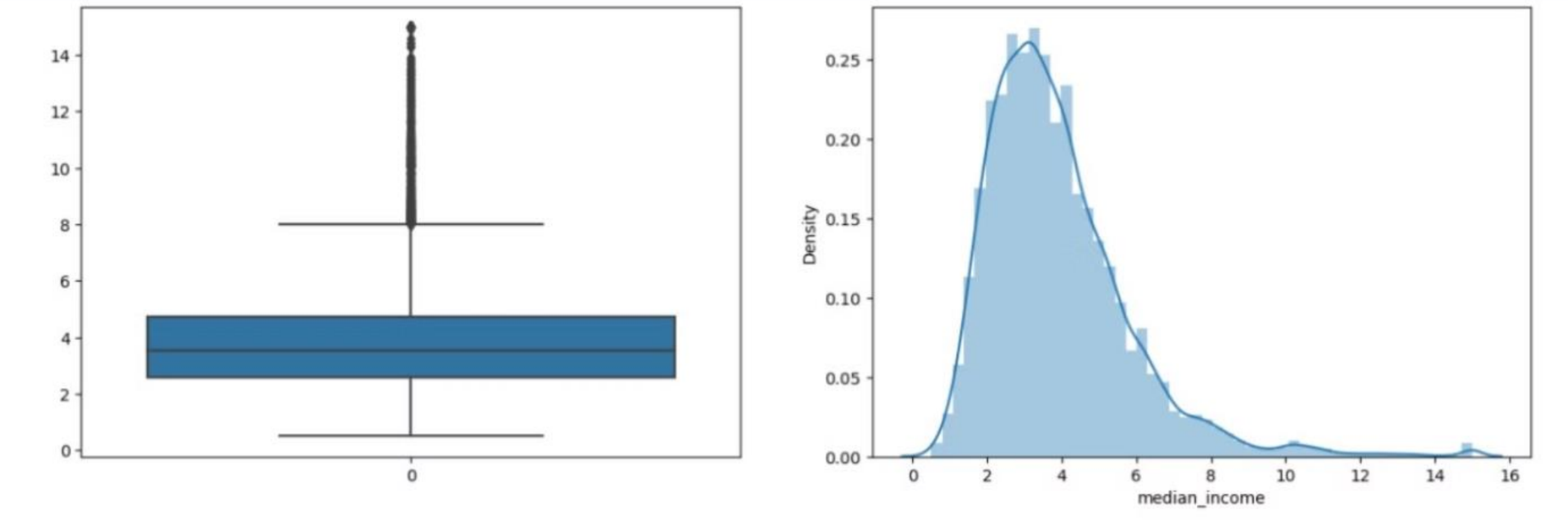
2. **Objective Function Optimization:** Recognizing limitations in estimating objective function values based on cluster centroids rather than actual pipeline placements.

3. **Computational Resources:** Managing computational resources, resources, especially for large datasets, when conducting cost cost estimations using Clustering and Regression algorithms. algorithms.

# Solution and Approach:

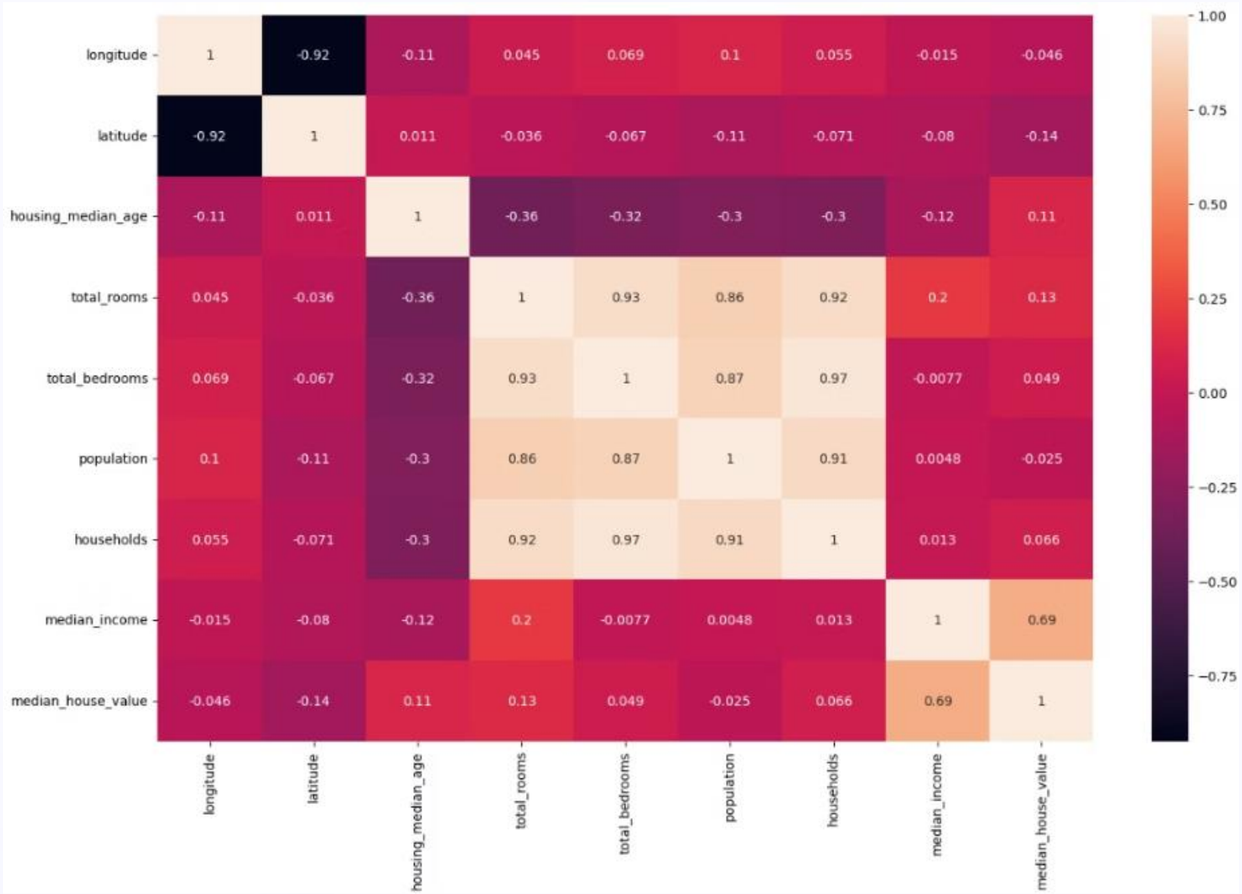
1. **Outlier Identification:**

- Initial step involved identifying outliers within the dataset through graphical analysis.
- Outliers could potentially skew clustering results or affect cost estimations.



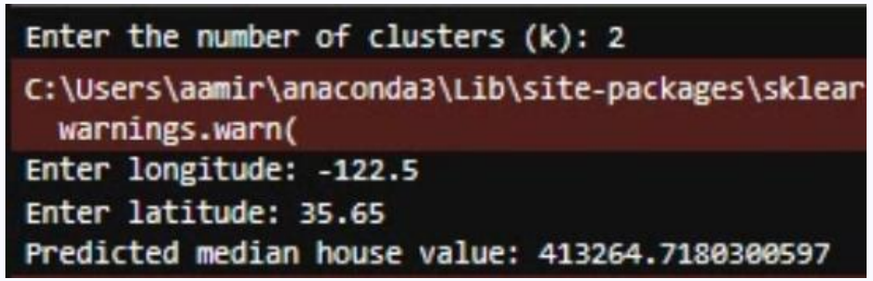
## Correlation Heatmap Analysis:

- Analyzed a correlation heatmap to understand the relationship primarily between longitude and latitude.
- Assisted in formulating clustering strategies and pipeline alignment decisions.



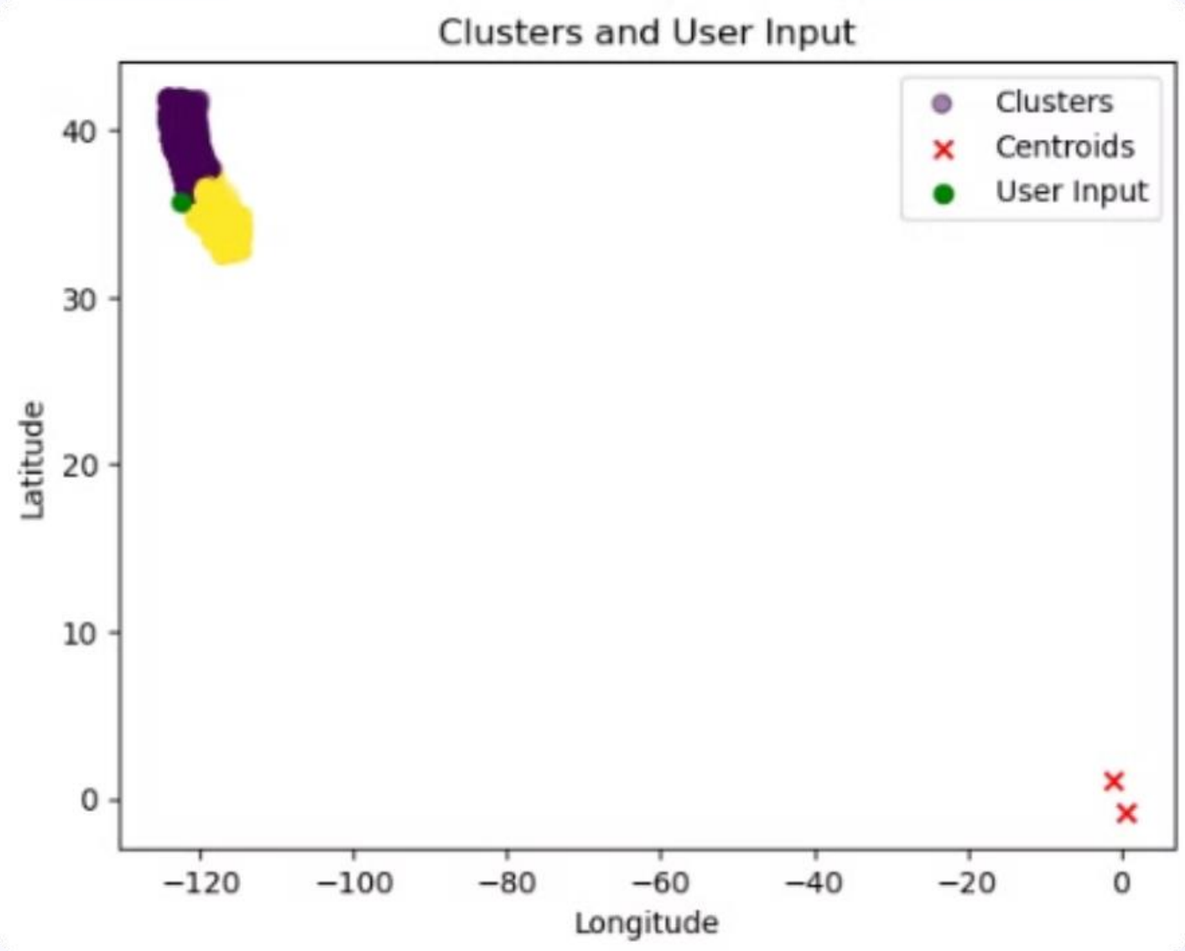
**Objective 1:** Efficient Line:

- Applied K-Means clustering to group houses into clusters for efficient pipeline routing.
- Clustering results provided insights into spatial distribution and potential pipeline alignments.
- Constructed a predictive model that predicts Median House Value by taking Longitude and Latitudes as input.
- Following is an output :



- Presence of outliers in the dataset cause the centroids to be pulled towards them. Causing them to be located outside the cluster.

- In higher-dimensional spaces, concept of a centroid becomes less intuitive, and may appear to be outside the cluster even if it is technically within the cluster dimensions.

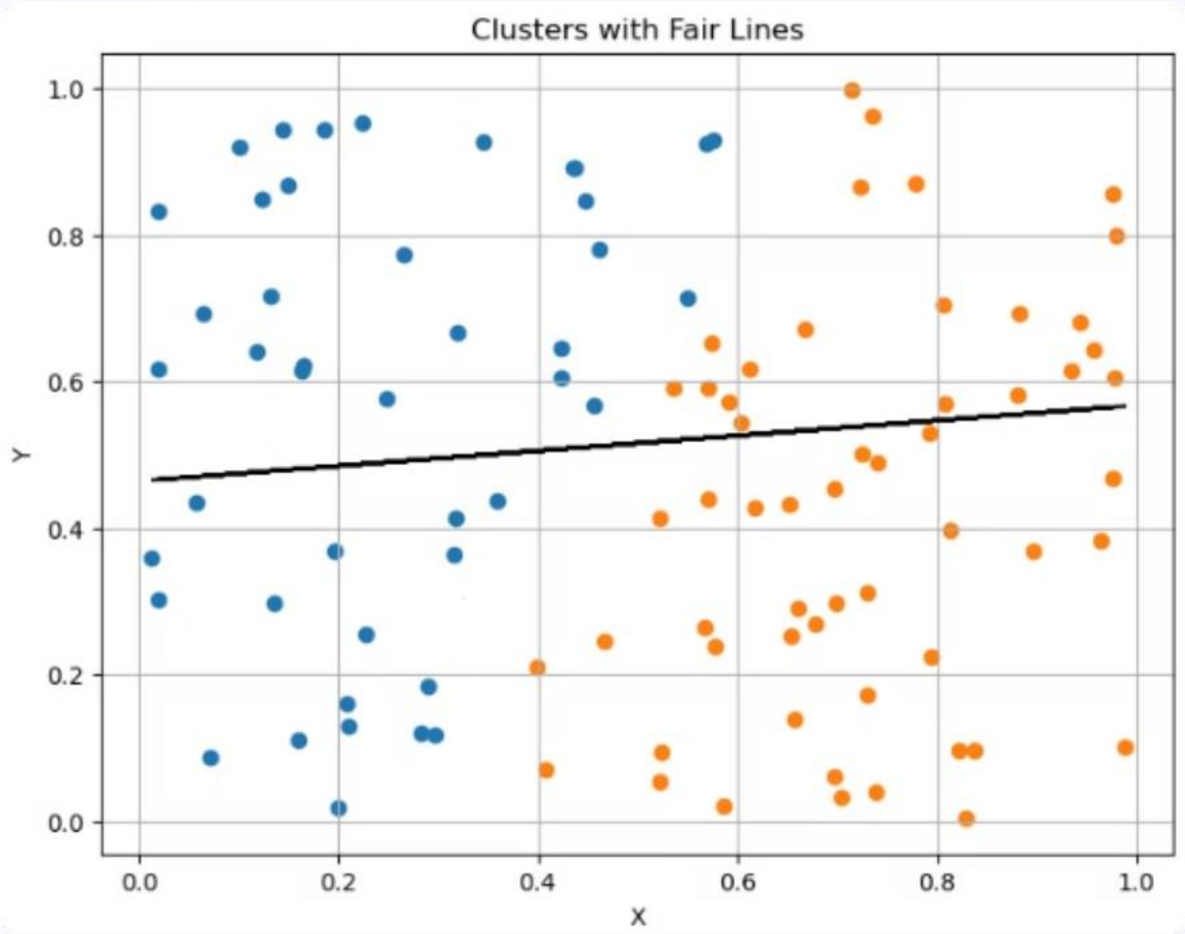


**Objective 2:** Fair Line:

- Employed K-Means clustering to determine centroids representing fair pipeline alignments.
- Utilized slope-intercept optimization to find the optimal line that minimizes the maximum distance to any house. (Each time the file is run arbitrary values are taken, resulting in different outputs) O/p :-

```
Optimal slope and intercept: [-3.36755043  0.06088739]
Minimized maximum distance to any house: 2.145561378154791
```

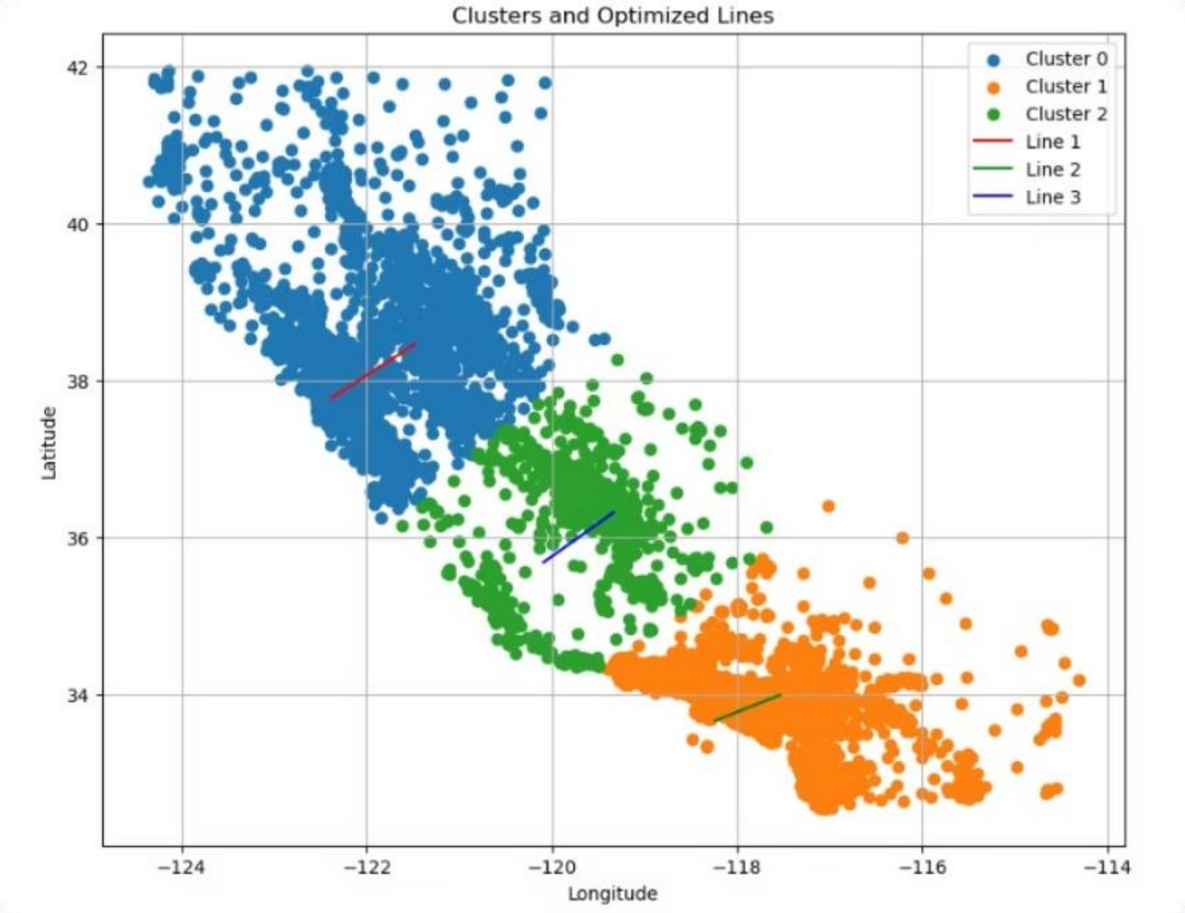
- Graphical representation facilitated the selection of pipeline routes minimizing maximum distance to any house.



**Objective 3 :** Multiple Efficient Lines:

- Utilized K-Means clustering to identify centroids representing potential pipeline alignments for multiple efficient lines.
- Calculated the minimum distance from each house to each centroid.
- Estimated the minimum cost of multiple efficient lines by summing the minimum distances from each house to the nearest centroid for each line.
- The output generated would be different each time as it takes the values arbitrarily. O/p :-

```
Efficient Lines:
Line 1: Direction = [0.90404439 0.69002502], Point = [-121.93029495   38.11365147]
Line 2: Direction = [0.69962205 0.3277204 ], Point = [-117.88493836   33.82435098]
Line 3: Direction = [0.75677864 0.63606106], Point = [-119.71000542   35.99619848]
Total Cost: 11511.599397055554
```



# Silhouette score

It is an extension of k means clustering algorithm where the values range from -1 to 1. And in which the value closer to 1 denotes that better clustering has been performed.

1. Objective 1 Silhouette score:

```
Silhouette Score: 0.7549172189258013
```

2. Objective 2 Silhouette score:

```
Silhouette Score: 0.7562423686755555
```

3. Objective 3 Silhouette score

```
Silhouette Score: 0.6471818705345762
```

# Conclusion:

- **Comprehensive Approach:** Our methodology integrated outlier identification, correlation analysis, and K-Means clustering to address the objectives efficiently.
- **Optimized Pipeline Routing:** By leveraging clustering techniques and optimization algorithms, we achieved efficient and fair pipeline routing solutions.
- **Cost-Effective Solutions:** Our approach enabled the estimation of pipeline pipeline costs while considering spatial distribution and fairness objectives. objectives.
- **Future Directions:** Further research could explore advanced clustering algorithms and optimization techniques for enhanced pipeline design and cost estimation.