

Telco  
Company

# Customer churn



# Analysis context

**12-15%**

churn rate in broader  
utilities sectors across  
Western Europe

**30-35%**

average churn rate for  
telecom companies

**Trello** Workspaces Recent Starred Templates Create

Customer Churn Telecoms Analysis ⚡ Workspace visible Board

Power-Ups Automation Filters & Share ...

**Data Collection**

- Flat file on Customer Churn
- Web Scraping
- BigQuery (Loading data)

+ Add a card

**Data cleaning and transformation**

- Cleaning csv flat file in python (2/2)
- ERD building and tables creation with SQL
- Cleaning web scraped data with SQL (1/1)

+ Add a card

**EDA and visualisation**

- Graph on categorical measures
- Graph on numerical measures
- Insights from graph
- Hypothesis testing to validate 3 insights
- SQL queries to analyse and compare metrics

+ Add a card

**Machine learning**

- Cluster analysis on customer churn in python
- Cluster interpretations and insights

+ Add a card

**Tableau creation**

- Summary dashboard (demographics, contracts, clusters)
- Service usage and tenure dashboard

+ Add a card

**Data exposition (API)**

- Flask app creation
- SQL code to expose data on flask
- Deployment in production

+ Add a card

## 1. Initial customer churn analysis

- Kaggle flat file
- EDA using Python
- Machine learning with clustering algorithm
- Tableau creation

## 2. Complementary data on real-life carriers

- Web-scraped data
- ERD creation
- EDA using SQL
- Loading data on BigQuery
- Api creation using flask

## Kaggle.com flat file

**Customer demographics**

**Account information**

**Services subscribed**

**Customer churn**

## Whistleout.com web-scraping

**Best phone plans per category**

**Best bundle plans per category**

**Carriers list and internet plans**

**T-Mobile family plans**

# Reducing churn rate



## Analysing

Analysing customers' profiles, accounts information, and services used.



## Clustering

Identifying common behaviors and profiles to then group customers into segments with similar characteristics.

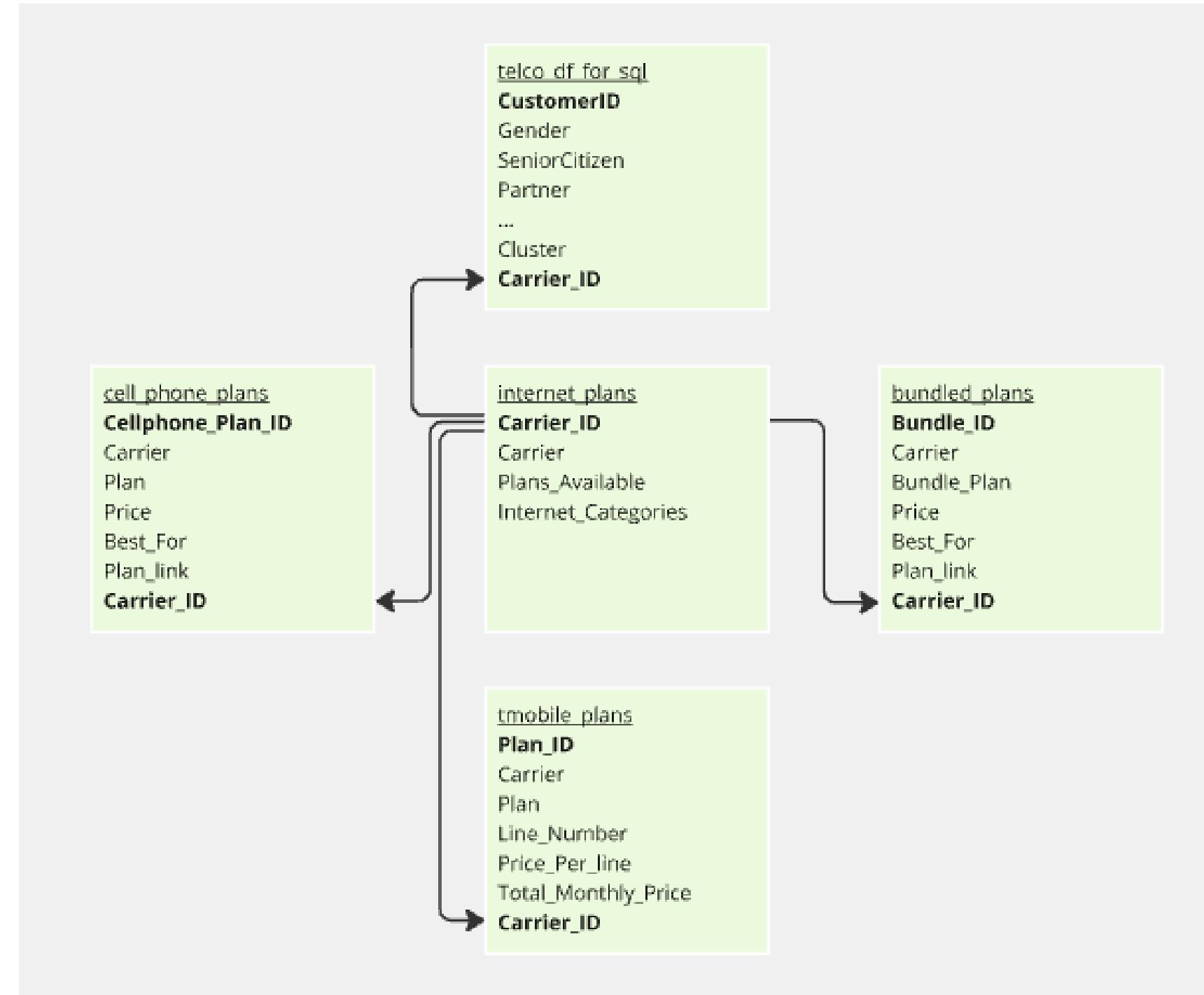


## Predicting

Identifying factors impacting the churn rate, and coming up with recommendations to reduce it.



# Entity Relationship Diagram



# Exploratory Data Analysis (EDA)

Customers are more likely to churn if:

**01** They have no dependent and no partner

**02** They are a senior citizen (older than 65y)

**03** They are recent customers (less than 1 year)

**04** They have multiple lines

**05** They have fiber optic Internet service

**06** Their contract is month-to-month

**07** They pay via electronic check

**08** They subscribe to Internet but do not subscribe to Online security, Online backup, Device protection, or Tech support services



# Exploratory Data Analysis (EDA)

Customers are more likely to churn if:

## 04 They have multiple lines

- T-Mobile family plans offer savings on multiple lines in order to reduce churn rate.
- Savings are even larger when analysing the total cost for 2 lines versus 3 lines in various phone plans.

```
SELECT
    Plan,
    Line_Number,
    Cleaned_Total_Monthly_Price,
    Cleaned_Total_Monthly_Price - LAG(Cleaned_Total_Monthly_Price) OVER (PARTITION BY Plan ORDER BY Cleaned_Line_Number)
    AS Price_Difference_Per_New_Line
```

```
FROM
    tmobile_plans
```

```
ORDER BY
    Plan,
    Line_Number;
```

Result Grid Filter Rows: Search Export:

Plan	Line_Number	Cleaned_Total_Monthly_Pri...	Price_Difference_Per_New_Line
► Essentials	2 lines	90.00	NULL
Essentials	3 lines	90.00	0.00
Essentials	4 lines	100.00	10.00
Essentials	5 lines	120.00	20.00
► Go5G	2 lines	140.00	NULL
Go5G	3 lines	140.00	0.00
Go5G	4 lines	155.00	15.00
Go5G	5 lines	180.00	25.00

# Hypothesis Testing

**Customers who have been using Telco's services for over one year are less likely to churn.**

Null Hypothesis ( $H_0$ ): Customers who have been using Telco's services for over one year are equally likely to churn as those who have been using it for less than or equal to one year.

Alternative Hypothesis( $H_1$ ):  
Customers who have been using Telco's services for over 1 year are less likely to churn.

**Chi-Square Statistic: 708.78**  
**P-Value: 0**

We reject the null hypothesis.

**Customers with month-to-month contracts are more likely to churn.**

Null Hypothesis ( $H_0$ ): Customers with month-to-month contracts are equally likely to churn as those with other types of contracts.

Alternative Hypothesis ( $H_1$ ):  
Customers with month-to-month contracts are more likely to churn.

**Chi-Square Statistic: 1153.97**  
**P-Value: 0**

We reject the null hypothesis.

**Senior citizens are more likely to churn than non-senior citizens.**

Null Hypothesis ( $H_0$ ): Senior citizens are equally likely to churn as non-senior citizens.

Alternative Hypothesis ( $H_1$ ): Senior citizens are more likely to churn than non-senior citizen

**Z-Statistic: 12.663**  
**P-Value: 0**

We reject the null hypothesis.

# Clustering Algorithm

## Step 1:

Transforming categorical features in numerical (dummification).

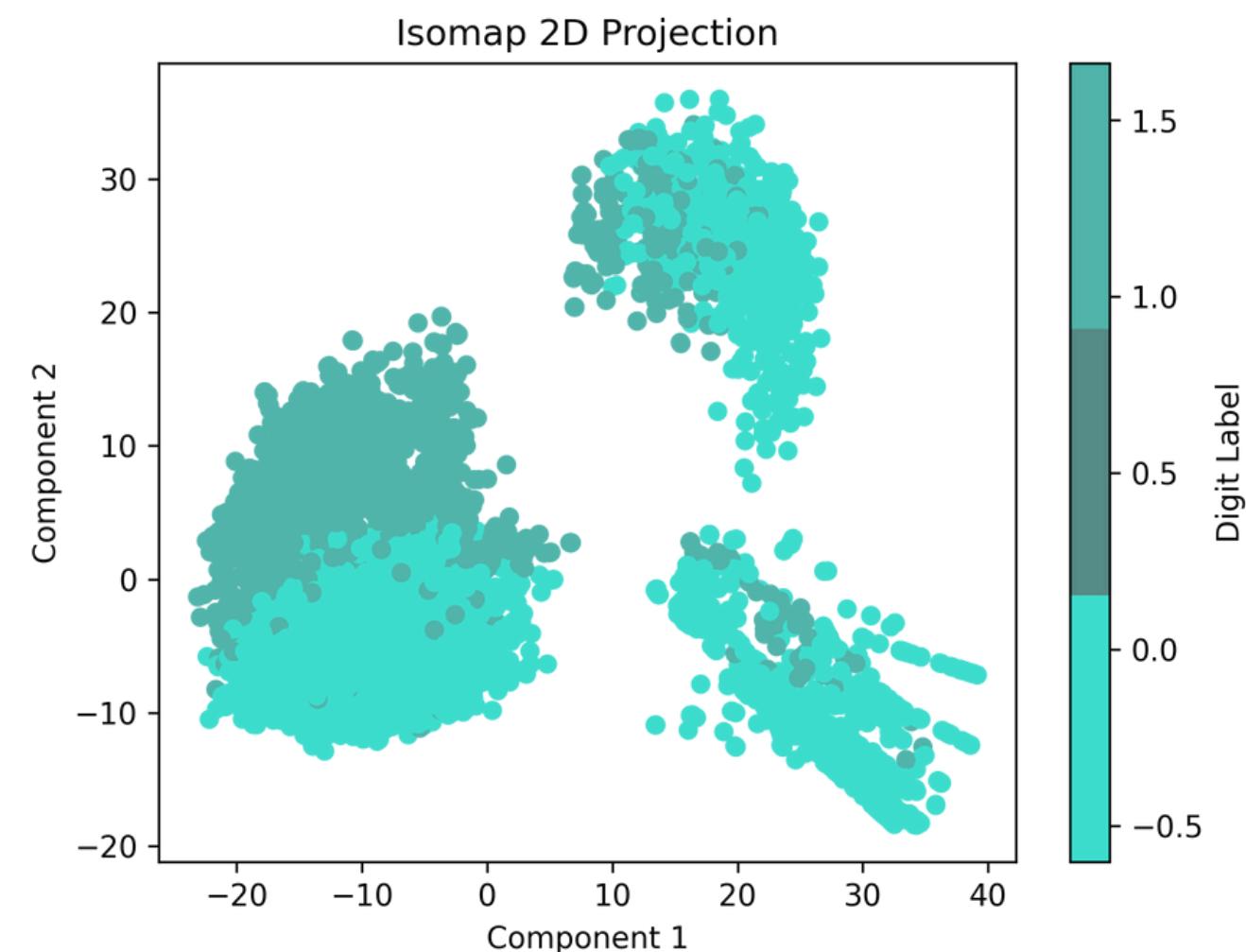
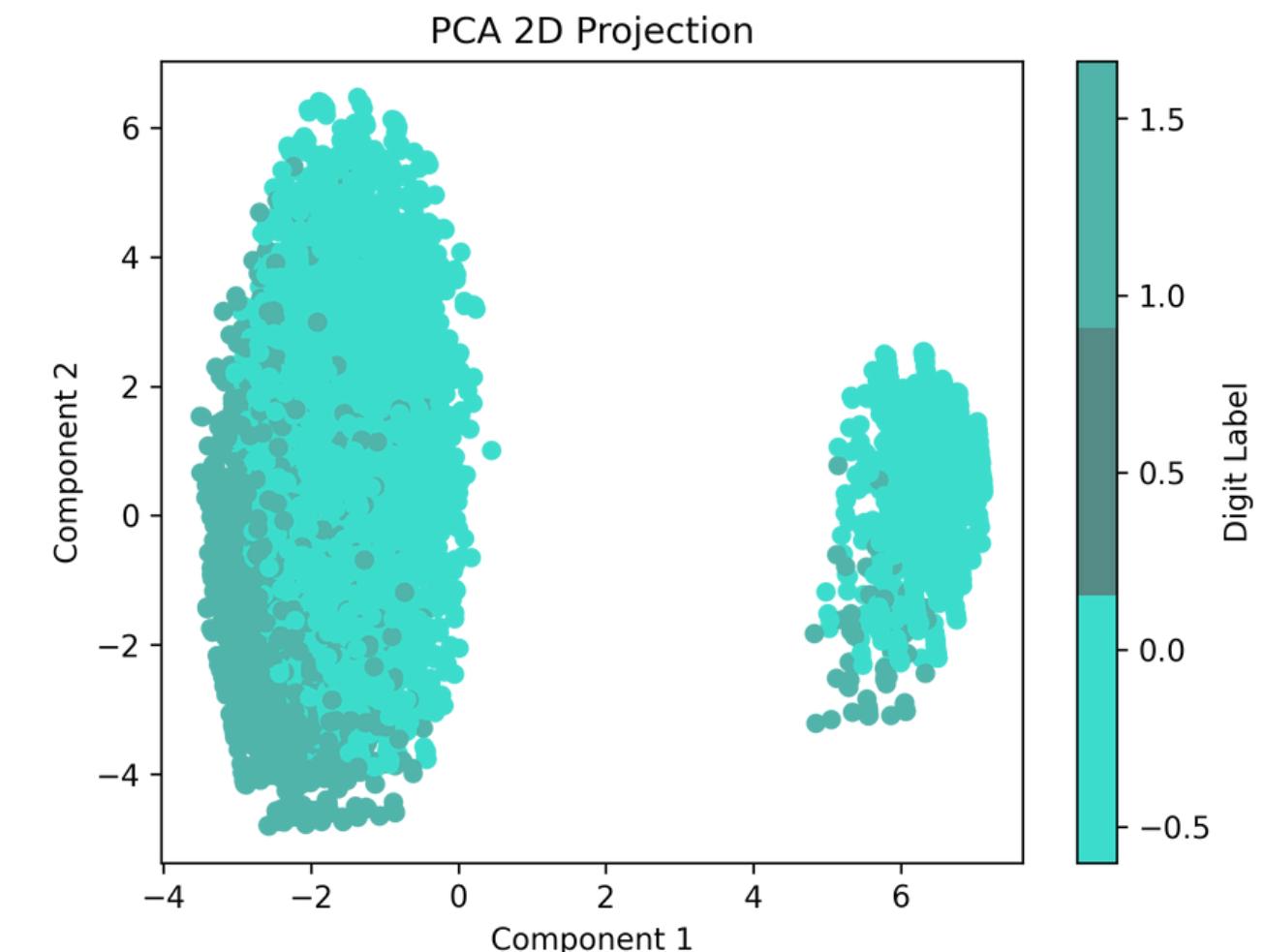
## Step 2:

Standardising the numerical data.

## Step 3:

Performing a dimensionality reduction method:

- PCA may not be optimal as the first two components explain only 0.38 of the variance, which is less than 80%.
- Isomap method seems to be more relevant than PCA for the 2D projection as we discovered a 3rd cluster.

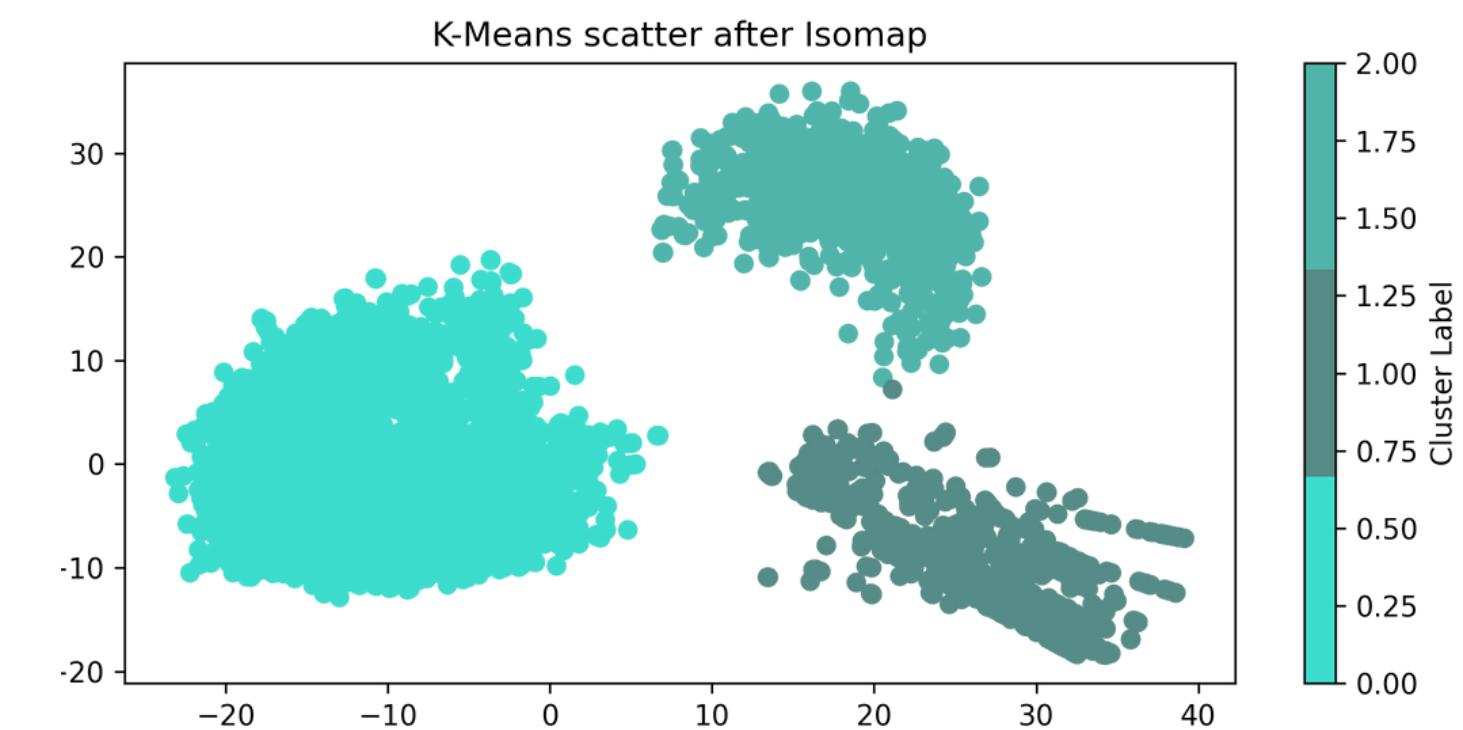
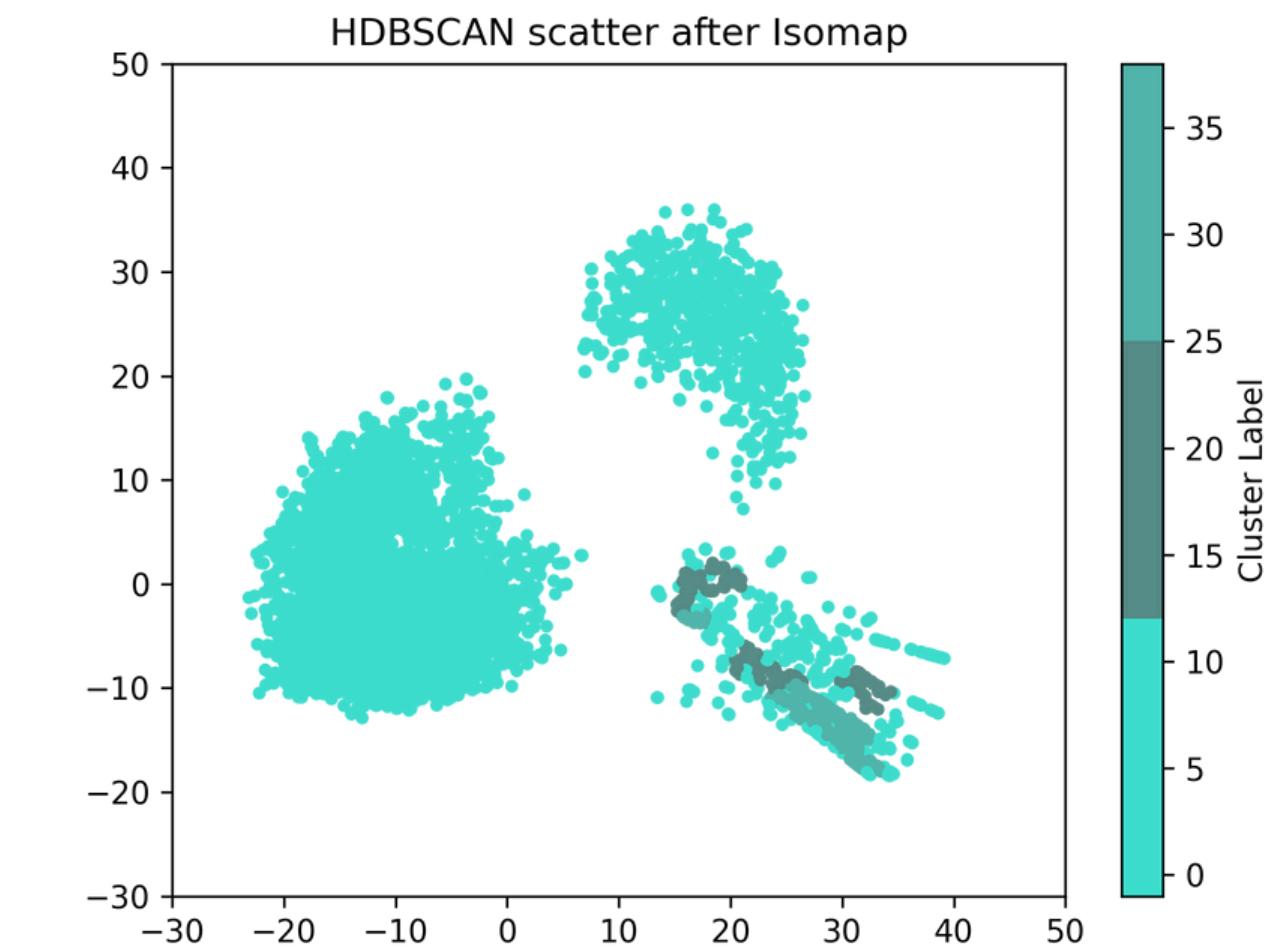
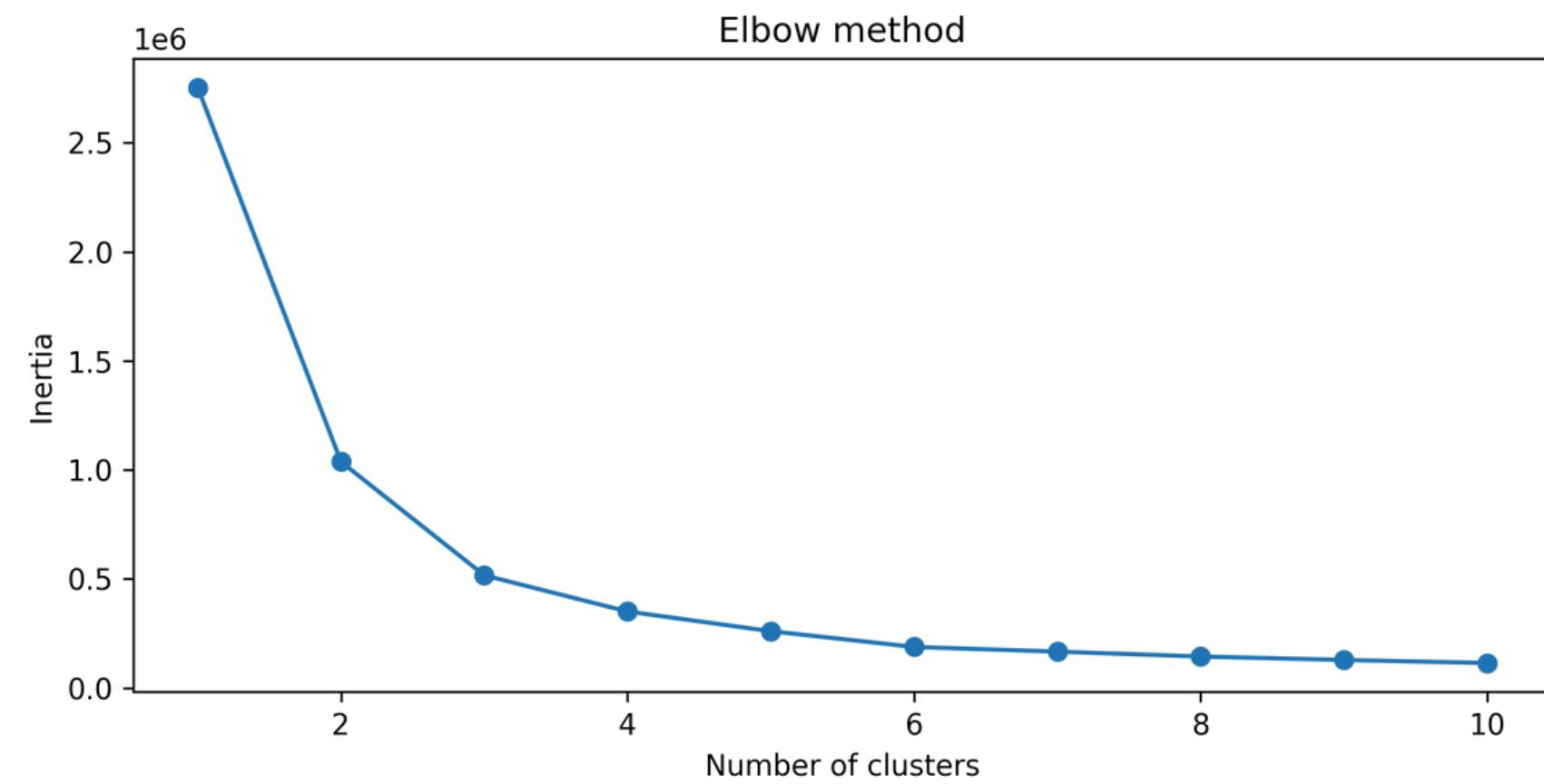


# Clustering Algorithm

## Step 4:

Clustering our data:

- K-means
- HDBSCAN



# Clusters

0

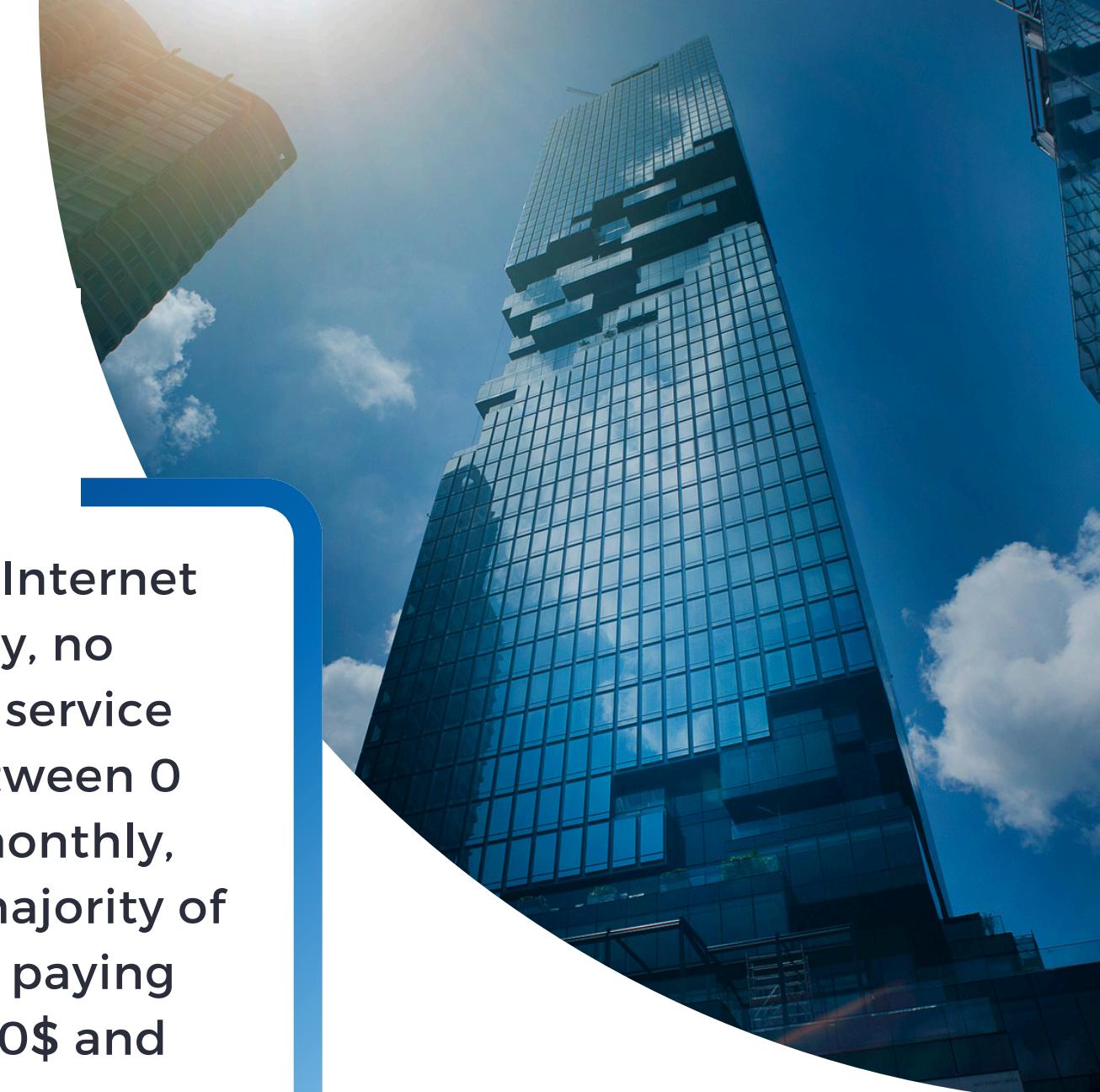
- Using both telephone and Internet services
- Mostly have no dependent
- Mostly have month-to-month contract, paperless billing and paying with electronic check
- Median monthly charge is 80\$
- 33% chance of churn

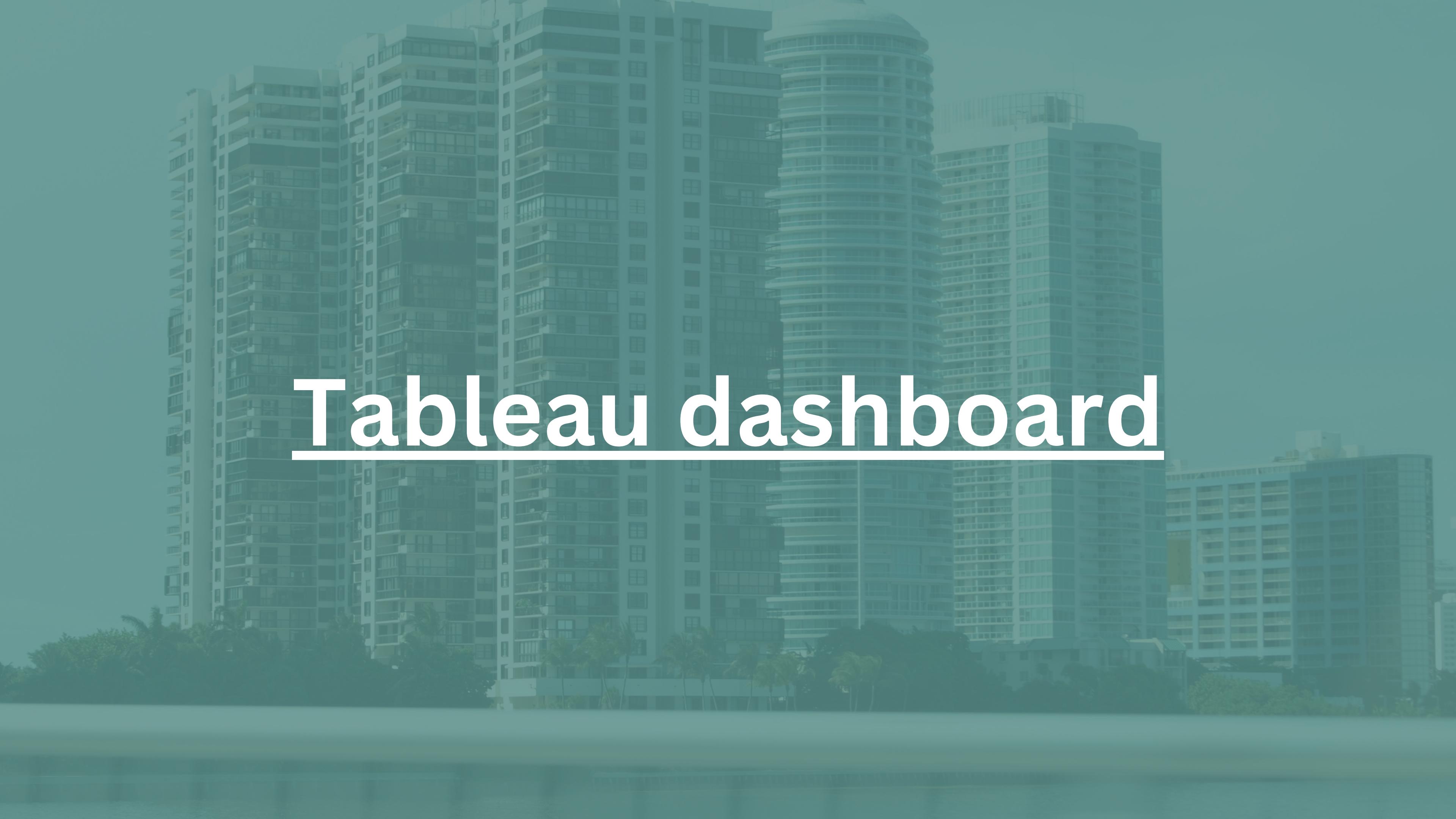
1

- Using telephone but no Internet service
- Mostly have 1 line
- Mostly under 65 years old
- Paying less than 30\$ monthly, mostly via mailed check
- Less than 10% chance of churn

2

- Using DSL Internet service only, no telephone service
- Paying between 0 and 80\$ monthly, with the majority of customers paying between 30\$ and 50\$.
- About 25% chance of churn





# Tableau dashboard

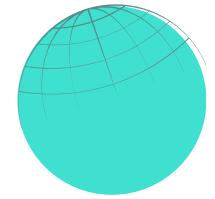
# Recommendations

- Churn prevention:
  - **All clusters:** Retain senior citizens with personalised offers.
  - **Cluster 0:** Retain customers with specific offers on additional internet service when using Fiber Optic, especially Online Backup and Device Protection. Run bundle offers to align with competitors. Intensify these offers in the first year of subscription.
  - **Cluster 2:** Run special offer on Tech Support.
- Loyalty program
  - **Cluster 1:** Reward customers to maintain engagement.
- Upselling
  - **All clusters:** Run incentive to subscribe to multiple telephone lines.
  - **Cluster 1:** Offer service upgrades, such as internet services.

# Improvements

---

- Additional measures like satisfaction score, customer lifetime value, or subscriptions' prices
- Information on competitors
- Set-up a system to match customers to the most appropriate plan



Telco  
Company

# Thank You!

Lucie Stenger  
Ironhack Bootcamp  
October 2024

# Appendix - Flask API

<http://localhost:8080/carrier/3>

```
{  
    "Carrier": "Spectrum",  
    "Carrier_ID": 3,  
    "Internet_Categories": "Cable",  
    "Plans_Available": 15  
}
```

[http://localhost:8080/Telco/cluster?  
page=0&page\\_size=10&cluster=1&gdpr=0](http://localhost:8080/Telco/cluster?page=0&page_size=10&cluster=1&gdpr=0)

```
{  
    "customers": [  
        {  
            "Carrier_ID": 40,  
            "Churn": "No",  
            "Cluster": 1,  
            "Contract": "One year",  
            "DeviceProtection": "No internet service",  
            "InternetService": "No",  
            "MonthlyCharges": 25.2,  
            "MonthlyCharges_cat": "0-30",  
            "MultipleLines": "Yes",  
            "OnlineBackup": "No internet service",  
            "OnlineSecurity": "No internet service",  
            "PaperlessBilling": "No",  
            "PaymentMethod": "Electronic check",  
            "TotalCharges": 1306.3,  
            "TotalCharges_cat": "1000-2000",  
            "tenure": 50,  
            "tenure_cat": "over 4 years"  
        },  
        {  
            "Carrier_ID": 40,  
            "Churn": "No",  
            "Cluster": 1,  
            "Contract": "Month-to-month",  
            "DeviceProtection": "No internet service",  
            "InternetService": "No",  
            "MonthlyCharges": 19.85,  
            "MonthlyCharges_cat": "0-30",  
            "MultipleLines": "No",  
            "OnlineBackup": "No internet service",  
            "OnlineSecurity": "No internet service",  
            "PaperlessBilling": "No",  
            "PaymentMethod": "Mailed check",  
            "TotalCharges": 57.2,  
            "TotalCharges_cat": "0-100",  
            "tenure": 3,  
            "tenure_cat": "0-1 year"  
        }  
    ]  
}
```

# **Appendix - GitHub repository**

**GitHub repository link**