

The background of the slide is a close-up photograph of a network switch or patch panel. Numerous Ethernet cables are plugged into the ports, with some having yellow and blue RJ45 connectors. The switch faceplate is dark grey with labels like '10/100Base-TX Ports (1-48)' and 'Link Mode'. Overlaid on the left side of the image is a large, abstract geometric shape composed of several overlapping planes in shades of blue and grey, creating a 3D effect.

Telecom Customer Churn Project

Project Overview



- Exploratory analysis of telecom customer dataset before predictive modeling.



- Objective: Understand customer demographics, data completeness, and churn-related attributes.



- Dataset: 7043 rows × 21 columns.



- Focus: Data exploration, cleaning, and structure understanding.

Objectives

- Load and inspect the dataset.
- Check missing or duplicate values.
- Analyze gender, partner, dependents, and senior citizen columns.
- Verify data consistency and quality.
- Prepare dataset for future visualization and modeling.

Importing Libraries

```
import numpy as np
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```



Purpose:

-
- numpy, pandas → data handling
-
- matplotlib, seaborn → visualization

Loading the Dataset

- `telecom = pd.read_csv('Telco-Customer-Churn.csv')`
- ✓ Loaded dataset into a pandas DataFrame for analysis.
- ✓ Verified data structure using `shape`, `info`, and `describe`.



Basic Exploration

- `telecom.head()` → View first 5 rows
- `telecom.tail()` → View last 5 rows
- `telecom.info()`, `telecom.describe()` → Data types and summary stats
-  Proper early data checks.
-  Could add `telecom.nunique()` for unique counts.

Column- Wise Details

- Used `telecom['gender'].reset_index()` etc.
- ⚠ Inefficient method.
- ✅ Better alternative:
- `telecom['gender'].value_counts()`
- → Easier and cleaner categorical analysis.

Filtering Examples

- `telecom.loc[telecom['Partner'] == 'No']` → 3641 without partner
- Gender: 3488 Female, 3555 Male
- Senior Citizen (60+): 1142
- Dependents: Yes 2110 / No 4933
-  Effective use of conditions.
-  Could simplify using `value_counts()`.

Missing & Duplicate Data

- `telecom.isnull().sum()` → No missing values
✓
- `telecom.count().duplicated()` ✗ Wrong
- ✓ Correct: `telecom.duplicated().sum()`
- → Checks for duplicate rows correctly.

Mistakes & Corrections

- Wrong: `telecom.count().duplicated()`
- Correct: `telecom.duplicated().sum()`
- Wrong: `.reset_index()` for viewing data
- Correct: `.value_counts()`
- No visualization included yet — should add plots for better understanding.

Key Findings

- Total customers: 7043
- Female: 3488
- Male: 3555
- Without Partner: 3641
- Senior Citizens: 1142
- Dependents: Yes 2110 / No 4933
- Missing: 0
- Duplicates: 0

Visualization Examples

- Gender Distribution → `sns.countplot(telecom['gender'])`
- Dependents Split → `plt.bar()`
- Senior Citizen Ratio → pie chart
- 💡 Graphs show clear demographic trends in telecom customer base.


Future Work

- **1** Visualize churn rates by gender, partner, dependents.
- **2** Handle categorical encoding for modeling.
- **3** Build churn prediction model (e.g., Logistic Regression).
- **4** Create dashboards for insights.

Conclusion

- ✓ Cleaned and validated telecom dataset.
- ✓ Checked missing and duplicate values.
- ✓ Identified key demographics.
- ✓ Corrected data exploration mistakes.
- ✓ Strong foundation for further churn prediction analysis.

Repository Structure

-  Telecom-Churn-Project/
- |
- |—— Telecom Project.ipynb
- |—— Telco-Customer-Churn.csv
- |—— README.md
- |—— images/
- |—— models/

Thank You

- Project by Chirag Jangid