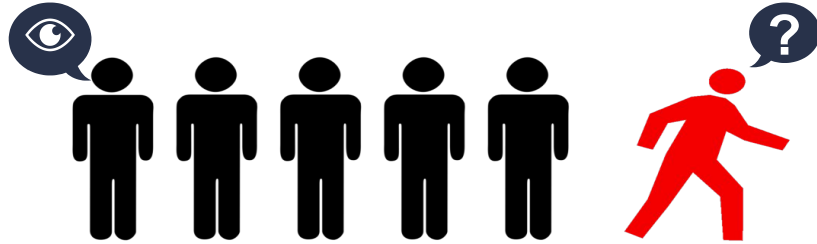


The background is white and features several decorative elements: a large teal circle with a white center in the top-left; a smaller teal circle below it; a large lime green circle in the top-right; a smaller green circle with a dashed outline below it; a large orange circle in the bottom-right; a smaller pink circle above it; a large green circle with a white center in the bottom-left; a smaller yellow circle above it; and a large yellow circle in the bottom-right. A dashed blue line curves from the top-left towards the bottom-right, passing behind the text.

Statistical Analysis

Using Python

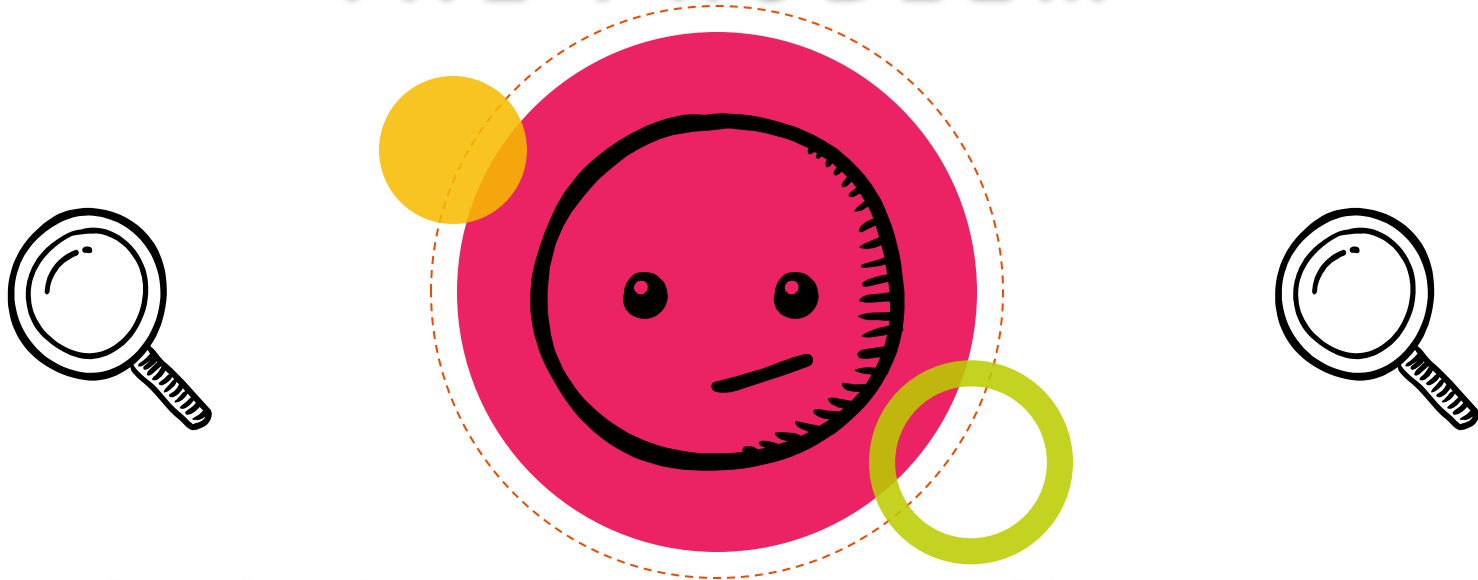
Telco Customer Churn



A telecommunications company is concerned about the number of customers leaving their land-line business for cable competitors.

They need to understand who is leaving.

UNDERSTANDING THE PROBLEM



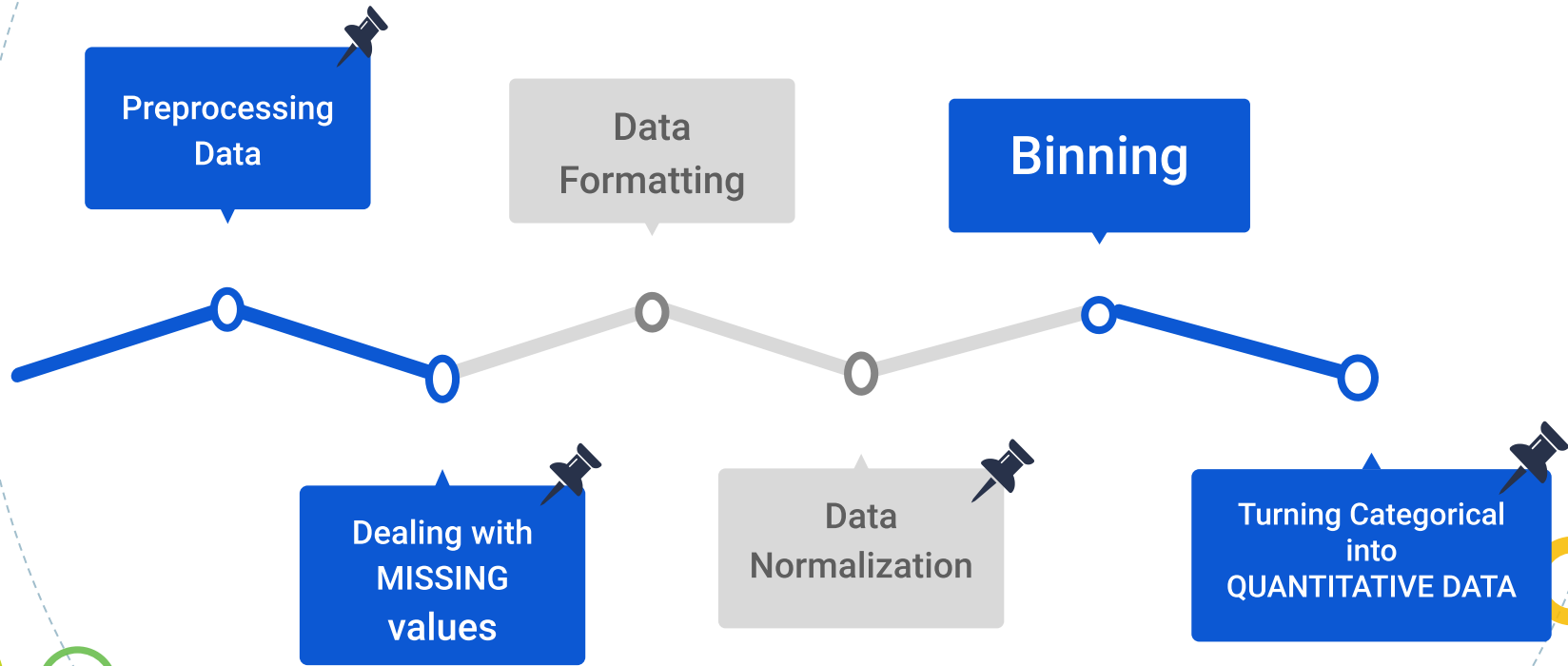
**Imagine that you're an analyst at this company
and you have to find out who is leaving and
why.**



About dataset

- **Customers who left within the last month** – the column is called Churn.
- **Services that each customer has signed up for** – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies.
- **Customer account information** – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges.
- **Demographic info about customers** – gender, age range, and if they have partners and dependents.

DATA WRANGLING



Exploratory Data Analysis

Descriptive Statistics

Giving short summaries about the sample and measures of Data.

ANOVA

Statistical comparison of groups.

Correlation

Measures to what extent different variables are Independent.



Model Development

Training Dataset

- The sample of data used to fit the model.
- The actual dataset that we use to train the model .
- The model sees and learns from this data.

Testing Dataset

- The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.
- The Test dataset provides the gold standard used to evaluate the model. It is only used once a model is completely trained.



Model Development

Logistic

Regression

It's a classification algorithm, that is used where the response variable is categorical.

The idea of Logistic Regression is to find a relationship between features and probability of particular outcome.

Model Evaluation

Logistic
Regression

Confusion
matrix

Log Loss

“

Thank You :)

Any Questions . . .