

EXPLORTORY DATA ANALYSIS

2.1 WHAT IS PRICE SENSITIVITY?

Price sensitivity is the degree to which demand changes when the cost of a product or service changes.

In the context of PowerCo, the “demand” refers to the demand for energy consumption.

Price sensitivity is commonly measured using the price elasticity of demand, which states that some consumers won’t pay more if a lower-priced option is available.

WHAT IS PRICE ELASTICITY OF DEMAND?

Price elasticity of demand is a measurement of the change in consumption of a product in relation to a change in its price.

2.2 QUESTION

1. Which of the following statements best describes price sensitivity?

- A. Price sensitivity is the difference between the maximum and minimum price in history
- B. Price sensitivity is how sensitive the price is at any given time
- C. Price sensitivity is the degree to which demand changes when the cost of product or service changes.

2. Which of the following statements best describes price elasticity of demand?

- A. A measurement of the change in consumption of a product in relation to a change in its price.
- B. A measurement of the change in price of a product in relation to the time history
- C. A measurement of the change in consumption of a person in relation to a change in a price.

2.3 EXPLORATORY DATA ANALYSIS

The client has sent over 2 datasets and it your responsibility to perform some exploratory data analysis.

WHAT IS EXPLORATORY DATA ANALYSIS?

Exploratory data analysis (EDA) is a technique used by a Data Scientist to gain a holistic understanding of the data that they are working with.

It is mainly based around using statistical techniques (such as descriptive statistics) and visualizations to gain a deeper understanding of the statistical properties that the data holds.

2.4 QUESTION

Which of the following is typically not a technique that a Data Scientist would use during exploratory data analysis?

- A. Data Visualiztion
- B. Deep Learning
- C. Descriptive Statistics

2.5 LET' GET FAMILIAR WITH THE DATA

As a Data Scientist at BCG, there will be occasions when you need to analyse data or investigate an issue and you are not provided strict instructions or guidance. You may be thinking, where do I start?

It is a highly valuable skill to begin to learn how to investigate a problem independently. A great way to learn this skill is to build a framework for analysis that work for you. In this step, you'll need to analyse client data sets using Python and upload your work as a Jupyter notebook. The client has sent over 3 data sets shown below:

1. Historical Customer data: Customer data such as usage, sign up date, forecasted usage etc
2. Historical pricing data: Variable and fixed pricing data etc
3. Churn indicator: Whether each customer has churned or not

You need to analyze the following using Python:

- The data types of each column
- Descriptive statistics of the dataset
- Distributions of columns

2.6 EXPLANATION

The first thing you should do is download the provided Jupyter notebook and the CSV datasets. To run the notebook, you need to make sure that you provide the path for the CSV files so that you can load the data. By running the cells that exist within the notebook from Estelle, this will show you what the two datasets look like, it will provide you with code to produce descriptive statistics and it will also give some examples and sample code on how to visualize the data.

Analysis – Once you've run the cells provided, it was your job to build on this exploratory analysis:

- The visualization provided by Estelle shows how many companies churned vs. how many companies did not churn. We can see from this that the churn rate is approximately 10%. This is actually a very good churn rate, the closer the rate is to 0%, the better.
- The next series of visualizations were created in an attempt to try and dive deeper into how churn changes based on other factors (using other columns). This is useful for us to investigate because it may help us to understand factors that drive churns.
- In the notebook we visualize churn vs. sales channel, forecast, number of years, consumption, and origin.

