Waze: Predicting User Churn

Hypothesis Testing

Project Overview

The objective of this project is to apply descriptive statistics and hypothesis testing in Python. This includes a two-sample hypothesis test to see if there is a statistically significant difference in the mean number of drives between drivers who use iPhones and those who use Android devices.

Key Insights

- The device variable was label encoded as **device_type**.
- The mean number of drives for iPhone users was **67.9** and for Android users it was **66.2**, for the whole dataset.
- A **sample of 20** was taken from the dataset for both iPhone and Android users.
- The sample mean of the iPhone sample was **80.85** and the sample mean from the Android sample was **61.7**.
- A two-sample t-test for the means was conducted, with a **t-score of 1.02** and a **P-Value of 0.31**.

Details

```
tstat, pvalue = stats.ttest_ind(a=iphone_drives,
                                b=android_drives,
                                equal_var=False)
print(f"t-test Statistic:", tstat)
print()
print(f'P-Value:', pvalue)
```

✓ 0.0s

```
t-test Statistic: 1.0275632679158548

P-Value: 0.311805450436491
```

The P-value is equal to 0.31 and so is greater than our significance level of 5% (0.05), therefore we fail to reject the null hypothesis. Thereby, there is no statistically significant difference between the mean number of drives between drivers who use iPhones and drivers who use Androids.

Next Steps

- Explore additional behavioural variables to identify more meaningful predictors of churn.
- Conduct further hypothesis tests across different user segments.
- Increase the sample size to strengthen statistical power.
- Transition to multivariate modelling to better explain churn behaviour.