Project description

Context

You are an analyst at a big online store. Together with the marketing department, you've compiled a list of hypotheses that may help boost revenue.
You need to prioritize these hypotheses, launch an A/B test, and analyze the results.

Description of the data

Data used in the first part of the project
`/datasets/hypotheses_us.csv` Download dataset

- `Hypotheses` — brief descriptions of the hypotheses
- `Reach` — user reach, on a scale of one to ten
- `Impact` — impact on users, on a scale of one to ten
- `Confidence` — confidence in the hypothesis, on a scale of one to ten
- `Effort` — the resources required to test a hypothesis, on a scale of one to ten. The higher the `Effort` value, the more resource-intensive the test.

Data used in the second part of the project
`/datasets/orders_us.csv` Download dataset

- `transactionId` — order identifier
- `visitorId` — identifier of the user who placed the order
- `date` — of the order
- `revenue` — from the order
- `group` — the A/B test group that the user belongs to

`/datasets/visits_us.csv` Download dataset

- `date` — date
- `group` — A/B test group
- `visits` — the number of visits on the date specified in the A/B test group specified

Make sure to preprocess the data. There might be mistakes in the original datasets; for example, some of the visitors might have gotten into both group A and group B.

Part 1. Prioritizing Hypotheses

The file `hypotheses_us.csv` contains nine hypotheses on boosting an online store's revenue with `Reach`, `Impact`, `Confidence`, and `Effort` specified for each. The task is to:

- Apply the `ICE` framework to prioritize hypotheses. Sort them in descending order of priority.
- Apply the `RICE` framework to prioritize hypotheses. Sort them in descending order of priority.
- Show how the prioritization of hypotheses changes when you use `RICE` instead of `ICE`. Provide an explanation for the changes.

Part 2. A/B Test Analysis

You carried out an A/B test and got the results described in the files `orders_us.csv` and `visits_us.csv`.

Task

Analyze the A/B test:

1. Graph cumulative revenue by group. Make conclusions and conjectures.
2. Graph cumulative average order size by group. Make conclusions and conjectures.
3. Graph the relative difference in cumulative average order size for group B compared with group A. Make conclusions and conjectures.
4. Calculate each group's conversion rate as the ratio of orders to the number of visits for each day. Plot the daily conversion rates of the two groups and describe the difference. Draw conclusions and make conjectures.
5. Plot a scatter chart of the number of orders per user. Make conclusions and conjectures.
6. Calculate the 95th and 99th percentiles for the number of orders per user. Define the point at which a data point becomes an anomaly.
7. Plot a scatter chart of order prices. Make conclusions and conjectures.
8. Calculate the 95th and 99th percentiles of order prices. Define the point at which a data point becomes an anomaly.
9. Find the statistical significance of the difference in conversion between the groups using the raw data. Make conclusions and conjectures.
10. Find the statistical significance of the difference in average order size between the groups using the raw data. Make conclusions and conjectures.

11. Find the statistical significance of the difference in conversion between the groups using the filtered data. Make conclusions and conjectures.
12. Find the statistical significance of the difference in average order size between the groups using the filtered data. Make conclusions and conjectures.
13. Make a decision based on the test results. The possible decisions are: 1. Stop the test, consider one of the groups the leader. 2. Stop the test, conclude that there is no difference between the groups. 3. Continue the test.

How will my project be assessed?

Your project will be assessed on these criteria. Read over them carefully before you start the project.
Here's what project reviewers look for when assessing your project:

- How you prepare the data for analysis
- How you prioritize hypotheses
- How you interpret the resulting graphs
- How you calculate statistical significance
- What conclusions you draw based on the A/B test results
- Whether you follow the project structure and keep the code tidy
- The conclusions you make
- Whether you leave comments at each step