Hi Associate Director,

I hope you find this email well.

After the discussion with my team, we agreed that there are a few various hypotheses that should be tested out.

Null hypothesis: Predict the customer churn rate.

Null hypothesis: Indicate to which customers the 20% discount should be offered.

Alternative hypothesis: To test if the high churn rate is driven by customer price sensitivity.

The Methodology:

Executive summary
Perform data collection: customer historical data, customer data, customer transaction data, and customer churn data.

Perform data wrangling: check for duplications, and null values, categorise groups and apply one-hot encoding to categorical variables

Perform exploratory data analysis using visualisations. Such as company_size by plan_type, industry by plan_type, and payment_method by company size and industry to provide extra insights to understand customer behaviour. Churn rate by month, quarterly and annual. Relational and transactional experience measurement. Identify at-risk customers.

Perform predictive analysis using classification models. The models we decided on for predictions are Logistic Regression, KNN, SVC, Decision Tree, and Random Forest.

The best model would be picked based on the tradeoff between the complexity, the explainability, and the accuracy of the models. We would subsequently dive deeper into the reasons and impact of price changes on the customer churn rate. The model would allow us to size the business impact of the client's proposed discounting strategy.

Thank you for your time. Please do not hesitate to approach me if there are any questions or additional ideas.

Kind regards,
Chia Zen Orchard