Data analytics approach

[

Agenda

1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

Introduction

Overview of Data Analytics Approach

- This analysis aims to leverage data to provide actionable insights that can drive decision-making.
- The approach is structured into three key phases: Data Exploration, Model Development, and Interpretation & Reporting.
- Each phase is designed to build upon the previous, ensuring a comprehensive understanding of the data and the development of a robust predictive model.

Data Exploration

Understanding the Data

- **Variable Characteristics**:
  - Analyze the distribution of variables, check for skewness towards certain demographics (e.g., a dataset skewed towards younger age brackets), and validate the fields in the dataset.
- **Data Limitations**:
  - Identify any limitations in the data and gather external data sources like ABS data to enrich the model. For example, geographic remoteness of postcodes might indicate the need for a bicycle for transportation.
- **Correlation Analysis**:
  - Explore interactions between variables and check for multicollinearity. For instance, there may be a correlation between age and tenure, as older individuals may have longer tenures.
- **Data Transformation**:
  - Ensure that the data is in the right format for analysis. This might involve adjusting data types, aggregating data to the appropriate level, or integrating external datasets.
- **Assumptions & Improvements**:
  - Document any assumptions made during data exploration, note the limitations, and outline potential improvements if additional time and resources were available.

Model Development

Building the Predictive Model

Hypothesis Creation:

- Formulate a hypothesis related to the business question. Use statistical testing to validate or refute this hypothesis.

Statistical Testing:

- Perform relevant statistical tests to determine the validity of your hypothesis.

Feature Engineering:

- Create calculated fields, such as converting DOB into age brackets or determining high-margin products based on list price and standard cost.

Model Performance:

- Evaluate the model using appropriate metrics like residual deviance, AIC, ROC curves, and R-Squared.

Documentation:

- Thoroughly document the model's performance, any assumptions made, and the limitations encountered during development.

Interpretation

Visualizing and Presenting Results

- **Significant Variables**:
  - Interpret the significant variables and coefficients from a business perspective, explaining how they impact the business issue at hand.
- **Compelling Storytelling**:
  - Use data to tell a compelling story that supports the business case. This should be both quantitative and qualitative, making the insights accessible and actionable.
- **Visualization**:
  - Present findings through visualizations such as graphs and charts to enhance understanding and impact.

Thank you