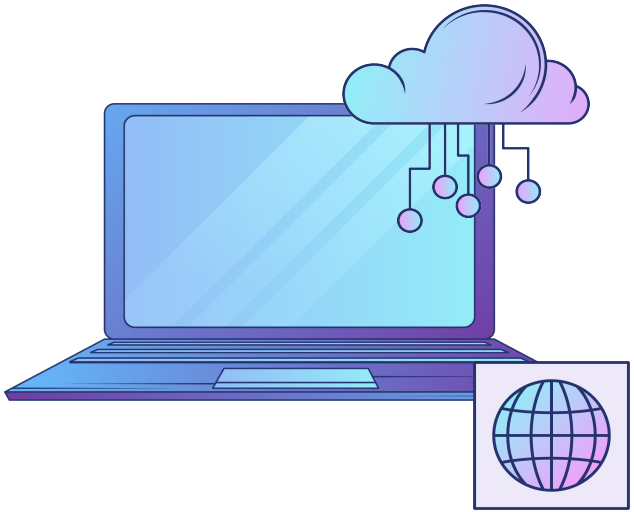
Click on Ads Prediction Analysis for Making Business Decisions

DECISION 520Q Section B Team 18
Yinan Chen, Tirth Pravin Gala, Yuhe (Tiffany) Jin, Chenjie (Angelina) Sun

Flow of Presentation

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Deployment
6. Key Takeaway



Business Understanding

1. **Industry & Challenge:**

   Focus: Telecommunication industry.

   Main Challenge: Not a well defined target segment of customers

2. **Impact:**

   Direct impact on revenue.

   Increases customer acquisition costs.

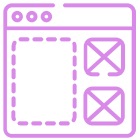
   Negative feedback from lost customers.

3. **Strategy:**

   Prioritize early identification of potential customers

   Recommendations as to refining our advertisement for better retention.

Objectives



Our aim

- Analyze potential customers in the telecom industry.
- Minimize cost of each impression and it's impact on revenue.
- Prioritize identification of patterns in viewers
- Recommend refinement in advertising strategies.

The tasks

- Predict customer retention or departures.
- Recommend fine tuning of retention strategies.
- Discover key product/service combinations.
- Offer tailored retention insights to optimize allocation of marketing resources

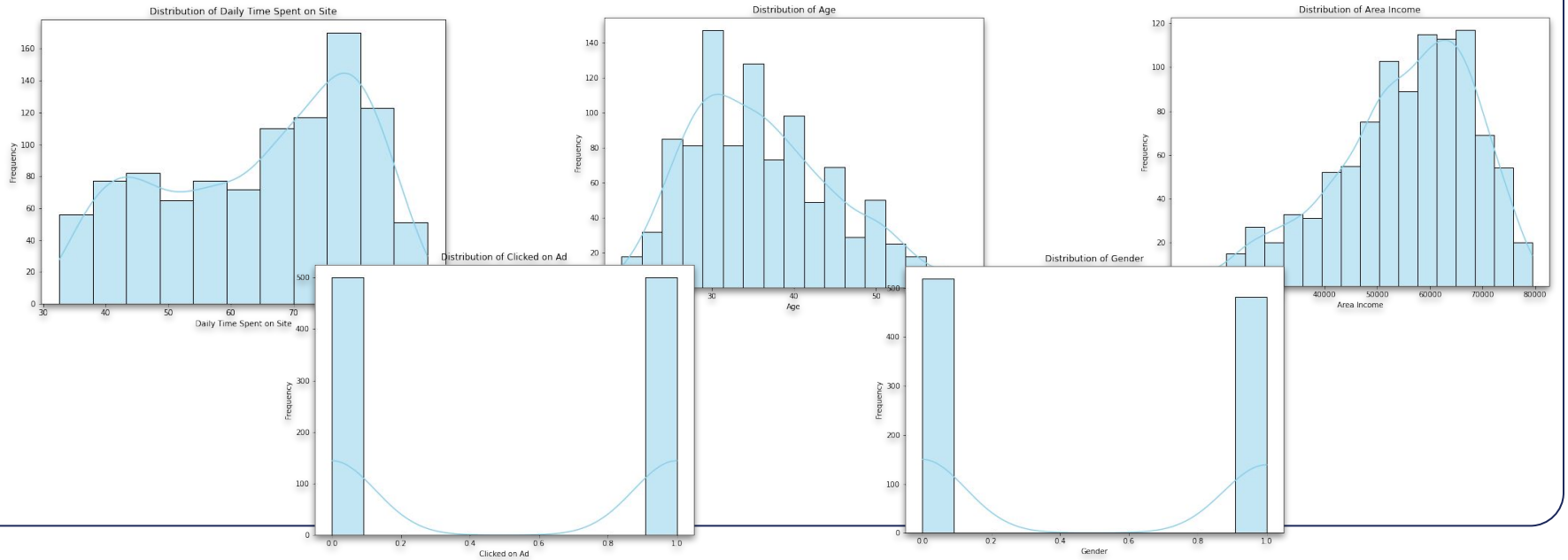
Data Understanding

- Data Source

  The ***Advertisement - Click on Ad dataset*** is retrieved from ***Kaggle***

- Exploratory Data Analysis

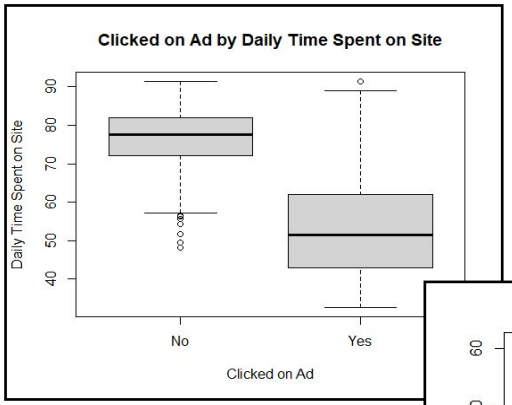
  5 Variables — Clicked On Ad, Daily Time Spent on Site, Age, Area Income, Gender

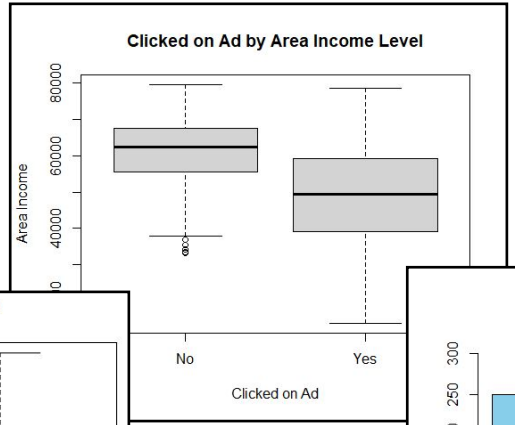


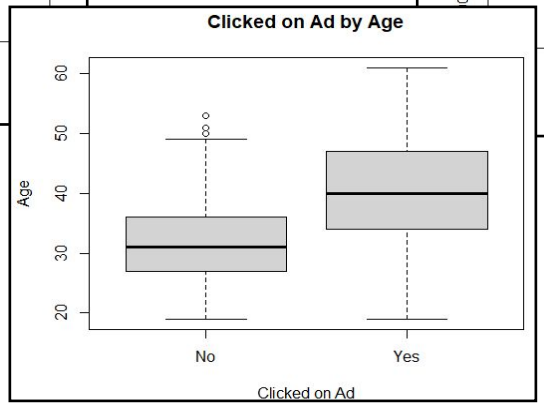
Data Understanding

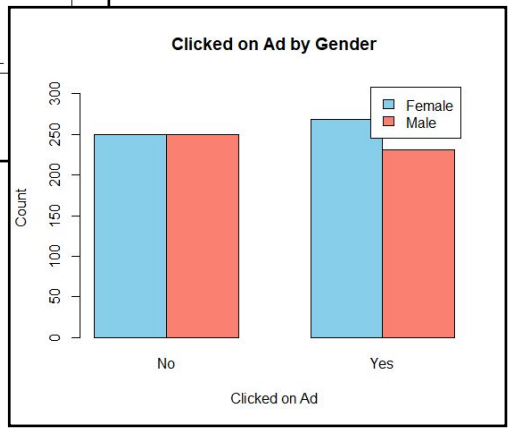
- Exploratory Data Analysis

  5 Variables — Clicked On Ad, Daily Time Spent on Site, Age, Area Income, Gender







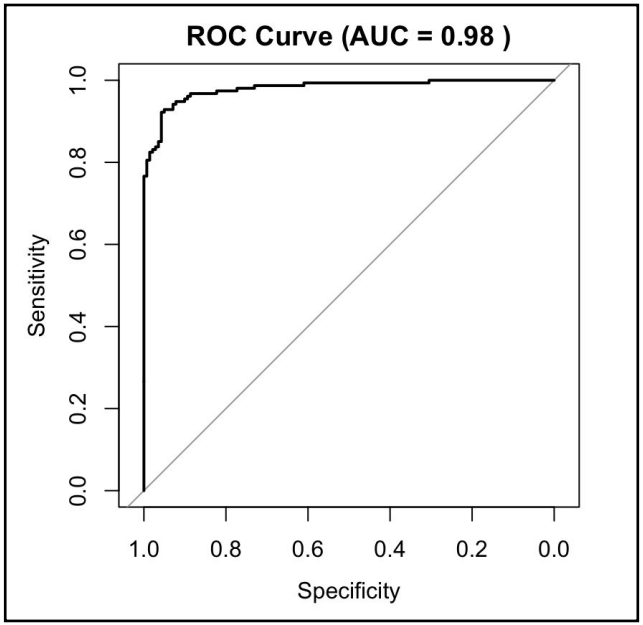


Modeling

5-Fold Cross Validation
Evaluation Metric: Accuracy

	In-Sample Accuracy	Out-of-Sample Accuracy
Logistic Regression	0.9206	0.9255
Decision Tree	0.7998	0.8315
CART	0.8374	0.8122
LASSO	0.9168	0.9205

ROC curve & AUC :
high discrimination ability.



Deployment

Deployment of Data Mining Result

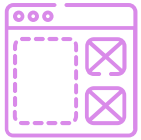
- Feed the data featuring characters of visitors of a certain website
- Predict a Clicked on Ad Rate for the exact ad
- Choose the websites where the ad can bring in the highest Clicks

Issues to be Aware of

- Biases and Inaccurate Results
- Model Robustness and Generalization
- Dramatic Change in the Market

Ethical Consideration

- Dataset itself may face some ethical issues
- Whether and where to place our ads would cause unexpected ethical problems



Other Risk & Proposed Solution

- Risk:
  specific features of a certain ad
- Solution:
  A/B testing on ad placement platforms

Proposed Solution and Insights

1

Focus on viewing behaviors of telecom ads on third-party sites.

2

Emphasis on total click count over click-through rate.

3

Tested models: logistic regression, decision tree, CART, LASSO.

4

Outstanding AUC of 0.98 on the ROC curve.

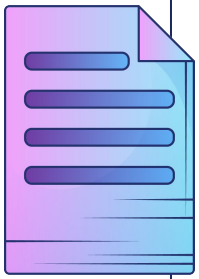
5

Lack of impression data led to focus on the raw click counts.

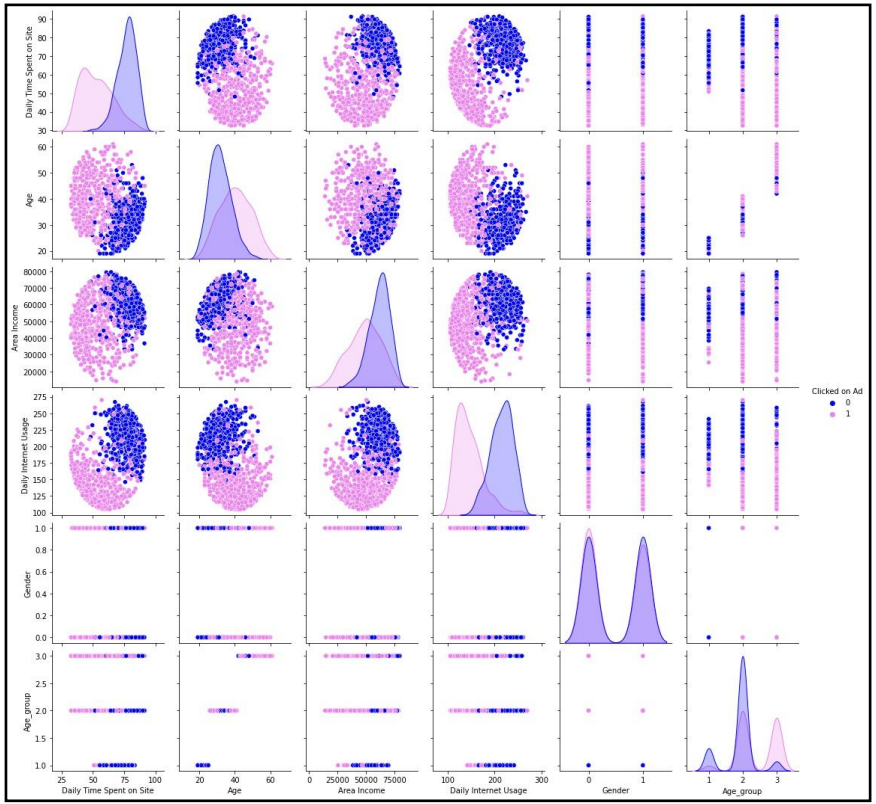
6

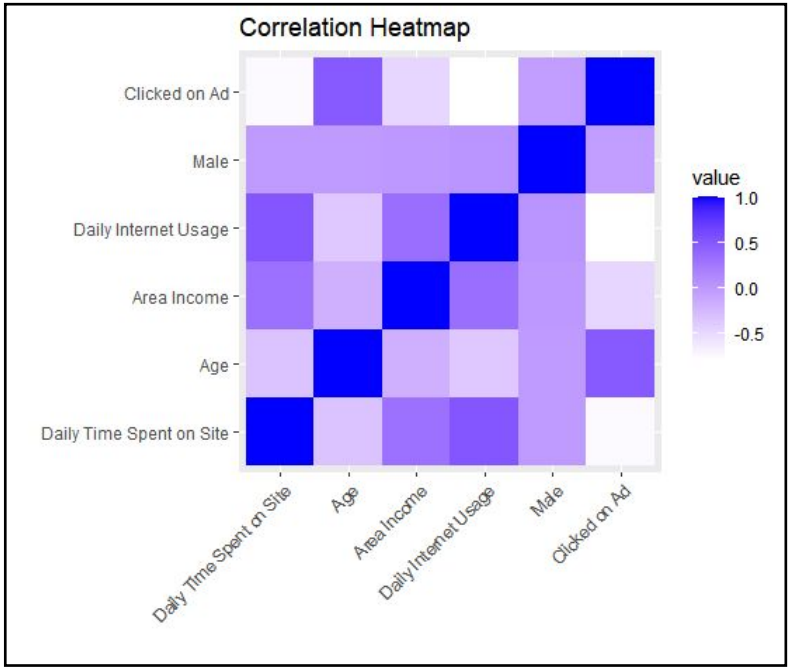
Analysis shows predictive modeling role in enhancing digital advertising.

Thank You!



Appendix


Daily Time Spent on Site
Age
Area Income
Daily Internet Usage
Gender
Age_group
Clicked on Ad
0
1


Correlation Heatmap
Clicked on Ad
Male
Daily Internet Usage
Area Income
Age
Daily Time Spent on Site
value
1.0
0.5
0.0
-0.5