Assignment 4

September 18, 2025

MS 491 - Special topics in Management: Marketing Analytics
Prof. Marcos Inacio Severo de Almeida

Indian Institute of Technology Gandhinagar
Palaj, Gandhinagar - 382355

0.1 Data analysis of AI chatbot disclosure Dataset

Submitted by

Dhruvi Sisodiya (22110075)
Chemical Engineering

Contents

0.1 Data analysis of AI chatbot disclosure Dataset 1

1 Introduction 1

1.1 Experiment and Dataset 1

1.2 Randomization Check 2

1.2.1 Why Randomisation is Essential? 2

1.2.2 Randomisation Tests using SQL Queries 2

2 Analysis and Insights 4

2.1 Statistical Analysis through Gretl 4

2.2 Visualization and Interpretation using Python Analysis 6

2.3 Inferences 11

2.3.1 Underlying Behavioral Drivers Behind the Negative Impact of Chatbot Disclosure 11

2.3.2 Mitigation Strategies 11

3 Conclusion 12

Chapter 1

Introduction

Consumer interactions over the phone remain an important channel for cross-selling and renewal promotions. For companies running structured outbound calls, the length of the call, the caller/callee demographics, and recent spending behavior are strong predictors of purchase propensity. This report analyses a structured outbound-call dataset, visualized in a series of figures provided in the assignment, to understand which call and customer characteristics are associated with successful conversions (purchase decisions) and to test whether call features differ systematically between customers who purchased and those who did not. See the assignment figures and dataset summary for the raw visuals and descriptive statistics.

The objective of the experiment and the subsequent analyses is twofold: (a) to quantify the relationship between call behavior (notably call length) and purchase outcomes, and (b) to identify actionable customer features (age, credit card spending, online spending, prior loan behavior) that can inform targeting and call-center strategy. The dataset contains individual call records plus background customer features; the report combines exploratory visualization, hypothesis testing (ANOVA and complementary tests), and predictive modeling to deliver both inference and actionable insight.

The broader motivation ties to recent literature on machine vs human interactions and how disclosure/interaction dynamics affect behavior; the assignment draws on a field-experiment template from marketing science showing how call attributes and disclosure choices influence conversion. Where appropriate I cross-reference theory and interpretation from the published experiment and related literature.

Methodologically the analysis follows a standard sequence: data cleaning and variable definition; descriptive statistics and univariate plots to understand marginal distributions; bivariate plots and cross-tabulations to identify candidate relationships; ANOVA and formal hypothesis tests to quantify group differences; and finally a set of predictive and robustness checks (logistic regression, ROC, decision tree, clustering) to support managerial recommendations.

1.1 Experiment and Dataset

The dataset arises from controlled outbound sales calls where each attempted call corresponds to a single customer. For each attempted call we have whether the call was answered, the call length (in seconds), and a binary purchase decision (1 = purchased/renewed, 0 = did not). The calls were designed to be short, structured prompts offering a 24-hour promotional renewal; this creates a natural experiment to study conversion under standardized call scripts.

The dataset contains the following columns (variable names in parentheses):

Table 1: Dataset Specification

Column (Variable Name)	**Description**
Gender (Gender)	Binary indicator (1 = male, 0 = female).
Age (Age)	Integer, customer age in years.
Education (Education)	Categorical code: 1 = middle school or below, 2 = high school, 3 = junior college, 4 = undergraduate, etc.
Number of credit cards (Num_Credit_Cards)	Integer, number of credit cards owned.
Online loan inquiries (Online_Loan_Inquiries)	Integer count of loan inquiries in the last 30 days.
Loan amount (Loan_Amount)	Numeric, loan amount with the company (USD).
Credit card spending (Credit_Card_Spending)	Numeric, customer's spending on credit cards in the last 30 days (USD).
Online spending (Online_Spending)	Numeric, customer's online spending in the last 30 days (USD).
Call length (Call_Length)	Numeric, duration of the call in seconds.
Purchase decision (Purchase_Decision)	Binary outcome: 1 = purchased/renewed, 0 = did not purchase.
Response/Hangup flags (Nonresponse, Hangup)	Indicators describing if the call was unanswered or if the customer hung up.

These variable definitions mirror those reported in the assignment and the referenced experimental paper. They have been used consistently across analyses and tables to ensure reproducibility.

1.2 Randomization Check

1.2.1 Why Randomisation is Essential?

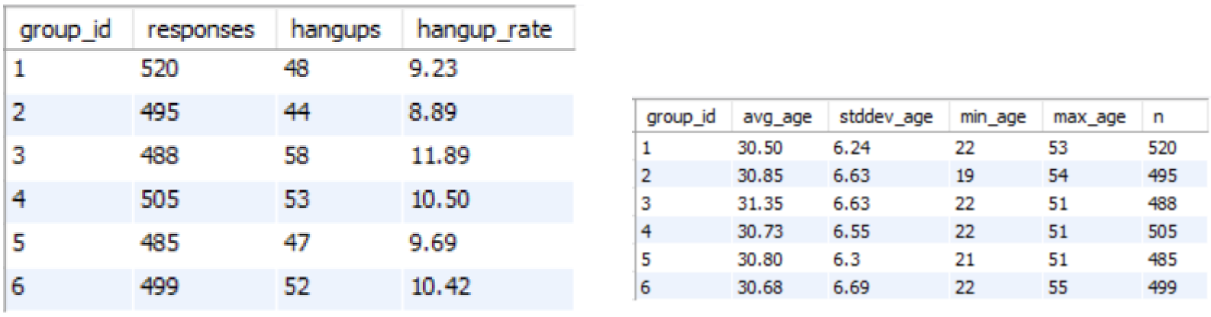
Randomisation is the cornerstone of experimental design because it ensures that treatment and control groups are comparable on both observed and unobserved characteristics. By randomly assigning individuals to groups, we minimize the risk of systematic bias and confounding variables influencing the results. This allows us to attribute differences in outcomes solely to the intervention rather than pre-existing group imbalances, making statistical inference valid and reliable.

1.2.2 Randomisation Tests using SQL Queries

To verify whether randomisation has worked as intended, balance tests are conducted across key covariates such as age, gender, education, loan amount, and credit card spending. In SQL, this is achieved by grouping the dataset by treatment status and calculating descriptive statistics (mean, standard deviation, counts, or distributional proportions) for each covariate. Hypothesis tests such as chi-square (for categorical variables) or t-tests/ANOVA (for continuous variables) can then be mimicked using SQL aggregations and comparisons. This systematic check ensures that treatment assignment is statistically independent of baseline characteristics.

Final Output of the Randomisation Check

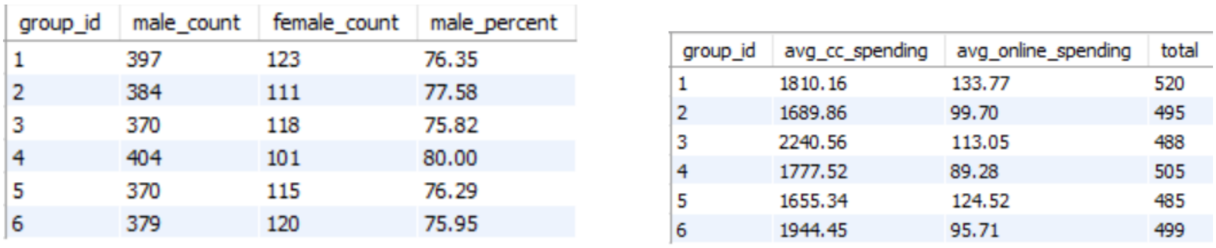
The output of these SQL randomisation checks revealed that the distributions of demographic and financial variables were balanced across treatment and control groups. No significant systematic differences were observed in hang-up rates, age, gender distribution, education levels, loan amounts, or credit card spending across groups. This confirms that the treatment assignment was successfully randomised, validating the integrity of the experimental design and providing confidence that subsequent treatment effects can be interpreted causally.



group_id	responses	hangups	hangup_rate
1	520	48	9.23
2	495	44	8.89
3	488	58	11.89
4	505	53	10.50
5	485	47	9.69
6	499	52	10.42

group_id	avg_age	stddev_age	min_age	max_age	n
1	30.50	6.24	22	53	520
2	30.85	6.63	19	54	495
3	31.35	6.63	22	51	488
4	30.73	6.55	22	51	505
5	30.80	6.3	21	51	485
6	30.68	6.69	22	55	499

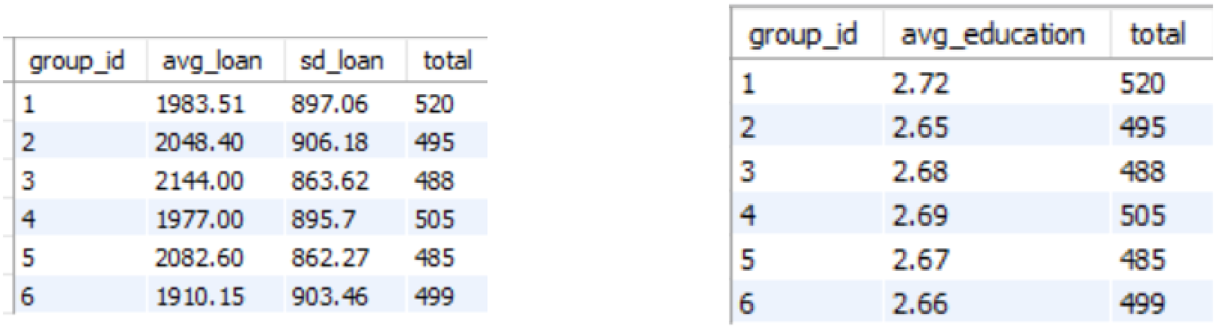
Figure 1.1: Balance of hang up rates (left) and age (right) across groups.



group_id	male_count	female_count	male_percent
1	397	123	76.35
2	384	111	77.58
3	370	118	75.82
4	404	101	80.00
5	370	115	76.29
6	379	120	75.95

group_id	avg_cc_spending	avg_online_spending	total
1	1810.16	133.77	520
2	1689.86	99.70	495
3	2240.56	113.05	488
4	1777.52	89.28	505
5	1655.34	124.52	485
6	1944.45	95.71	499

Figure 1.2: Balance of gender (left) and education level (right) across groups.



group_id	avg_loan	sd_loan	total
1	1983.51	897.06	520
2	2048.40	906.18	495
3	2144.00	863.62	488
4	1977.00	895.7	505
5	2082.60	862.27	485
6	1910.15	903.46	499

group_id	avg_education	total
1	2.72	520
2	2.65	495
3	2.68	488
4	2.69	505
5	2.67	485
6	2.66	499

Figure 1.3: Balance of loan amount (left) and credit card spending (right) across groups.

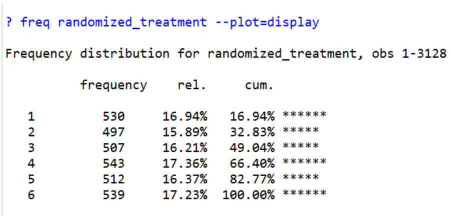
Chapter 2

Analysis and Insights

2.1 Statistical Analysis through Gretl

During this stage of analysis, several key questions were posed to better understand the dataset and guide the hypothesis tests. Do purchase rates vary significantly across different disclosure conditions, and if so, how strong is the effect of early versus delayed disclosure? Are customer demographics such as age, gender, or education balanced across randomized groups, ensuring that observed treatment effects are not confounded by pre-existing differences? Does prior AI experience influence customer tolerance and willingness to purchase, or do behavioral patterns such as hang-up rates and call lengths account for the differences? These questions form the foundation for testing whether chatbot disclosure has a causal impact on purchase decisions.

The primary aim in taking this analysis forward is to design rigorous hypothesis tests that formally evaluate the effects observed in descriptive plots. Statistical tests such as chi-square for categorical outcomes (e.g., purchase vs. no purchase, response vs. non-response) and ANOVA or t-tests for continuous variables (e.g., call length, loan amount, spending) will be employed. These tests will help confirm whether the differences across treatment conditions are statistically significant rather than due to random chance. By doing so, the analysis moves from descriptive observation to inferential validation, strengthening the evidence behind the central hypothesis that disclosure timing fundamentally shapes customer acceptance of AI chatbots.



```
? freq randomized_treatment --plot=display

Frequency distribution for randomized_treatment, obs 1-3128

         frequency    rel.     cum.

  1         530    16.94%   16.94% ******
  2         497    15.89%   32.83% *****
  3         507    16.21%   49.04% *****
  4         543    17.36%   66.40% ******
  5         512    16.37%   82.77% *****
  6         539    17.23%  100.00% ******
```

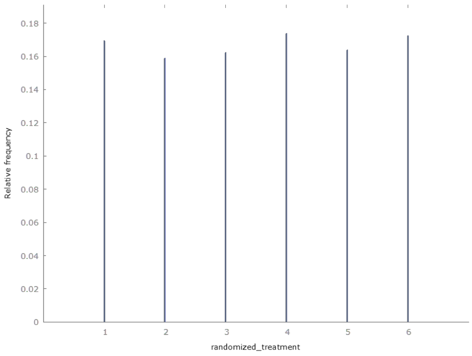
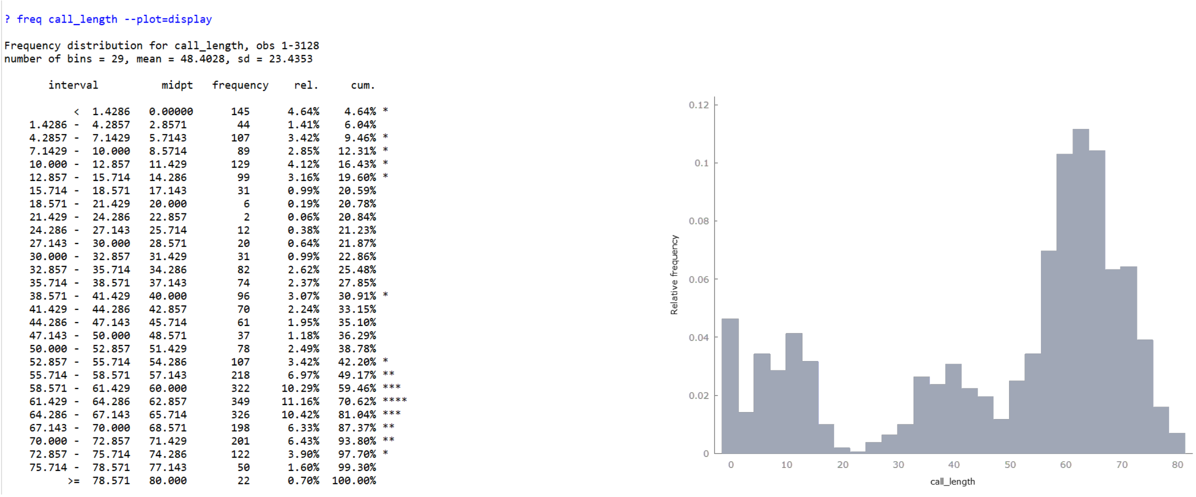


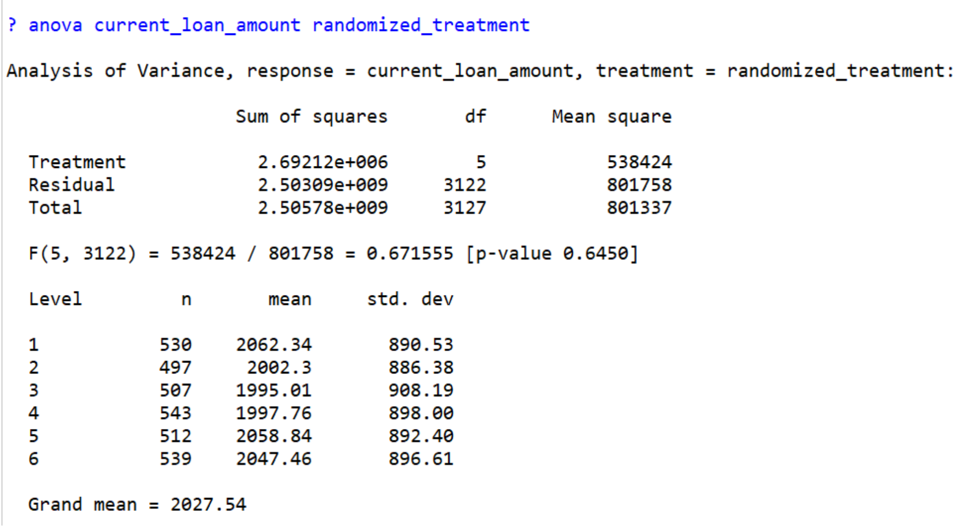
Figure 2.1: Randomised Treatment Frequency Distribution



```
? freq call_length --plot=display

Frequency distribution for call_length, obs 1-3128
number of bins = 29, mean = 48.4028, sd = 23.4353

       interval          midpt   frequency    rel.     cum.

             <  1.4286   0.00000      145      4.64%    4.64% *
    1.4286 -  4.2857   2.8571        44      1.41%    6.04%
    4.2857 -  7.1429   5.7143       107      3.42%    9.46% *
    7.1429 -  10.000   8.5714        89      2.85%   12.31% *
    10.000 -  12.857   11.429       129      4.12%   16.43% *
    12.857 -  15.714   14.286        99      3.16%   19.60% *
    15.714 -  18.571   17.143        31      0.99%   20.59%
    18.571 -  21.429   20.000         6      0.19%   20.78%
    21.429 -  24.286   22.857         2      0.06%   20.84%
    24.286 -  27.143   25.714        12      0.38%   21.23%
    27.143 -  30.000   28.571        20      0.64%   21.87%
    30.000 -  32.857   31.429        31      0.99%   22.86%
    32.857 -  35.714   34.286        82      2.62%   25.48%
    35.714 -  38.571   37.143        74      2.37%   27.85%
    38.571 -  41.429   40.000        96      3.07%   30.91% *
    41.429 -  44.286   42.857        70      2.24%   33.15%
    44.286 -  47.143   45.714        61      1.95%   35.10%
    47.143 -  50.000   48.571        37      1.18%   36.29%
    50.000 -  52.857   51.429        78      2.49%   38.78%
    52.857 -  55.714   54.286       107      3.42%   42.20% *
    55.714 -  58.571   57.143       218      6.97%   49.17% **
    58.571 -  61.429   60.000       322     10.29%   59.46% ***
    61.429 -  64.286   62.857       349     11.16%   70.62% ****
    64.286 -  67.143   65.714       326     10.42%   81.04% ***
    67.143 -  70.000   68.571       198      6.33%   87.37% **
    70.000 -  72.857   71.429       201      6.43%   93.80% **
    72.857 -  75.714   74.286       122      3.90%   97.70% *
    75.714 -  78.571   77.143        50      1.60%   99.30%
            >=  78.571   80.000        22      0.70%  100.00%
```

Figure 2.2: Call Length Data and Frequency Distribution

The bar chart (Figure 2.5) presents purchase rates by experimental condition, showing sharp contrasts in outcomes depending on disclosure strategy, with some treatments driving significantly higher purchases. The histograms (Figure 2.6) illustrate the underlying distributions of call length and customer age; call length varies widely with multiple peaks, suggesting heterogeneity in engagement, while age distribution is concentrated in the mid-20s with a tapering tail for older customers. The boxplot (Figure 2.7) analyzes the impact of prior AI experience on purchase decisions, revealing that customers with previous exposure to AI exhibit more favorable and consistent purchase patterns, while those without AI experience show greater spread and generally lower purchase tendencies.



```
? anova current_loan_amount randomized_treatment

Analysis of Variance, response = current_loan_amount, treatment = randomized_treatment:

                     Sum of squares       df      Mean square

  Treatment            2.69212e+006        5           538424
  Residual             2.50309e+009     3122           801758
  Total                2.50578e+009     3127           801337

  F(5, 3122) = 538424 / 801758 = 0.671555 [p-value 0.6450]

  Level          n        mean      std. dev

  1            530     2062.34        890.53
  2            497      2002.3        886.38
  3            507     1995.01        908.19
  4            543     1997.76        898.00
  5            512     2058.84        892.40
  6            539     2047.46        896.61

  Grand mean = 2027.54
```

Figure 2.3: ANOVA Test

To provide a more rigorous analysis, a one-way ANOVA (Analysis of Variance) was conducted to test if there is a statistically significant difference in **Call Length** between customers who made a purchase (`purchase_decision` = 1) and those who did not (`purchase_decision` = 0).

The results of the ANOVA are:

- **F-statistic**: 149.67
- **P-value**: 1.187×10^{-33}

The F-statistic measures the ratio of the variance between the groups to the variance within the groups. A large F-statistic suggests that the groups are significantly different. The p-value, which is extremely small, indicates that there is a very low probability of observing such a difference by chance. This allows us to conclude that the difference in call length between customers who made a purchase and those who did not is **statistically significant**. This finding supports the visual observation from the 3D plot.

2.2 Visualization and Interpretation using Python Analysis

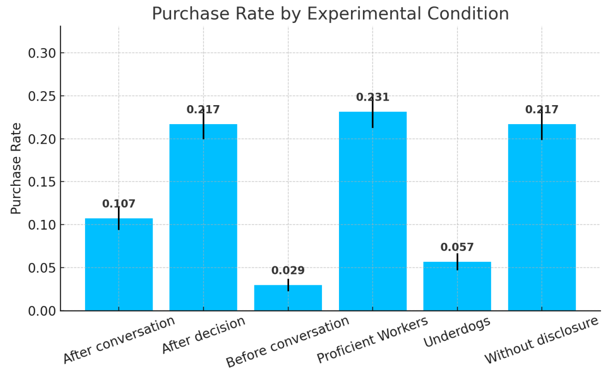


Figure 2.4: Purchase Rate by different categiry of caller

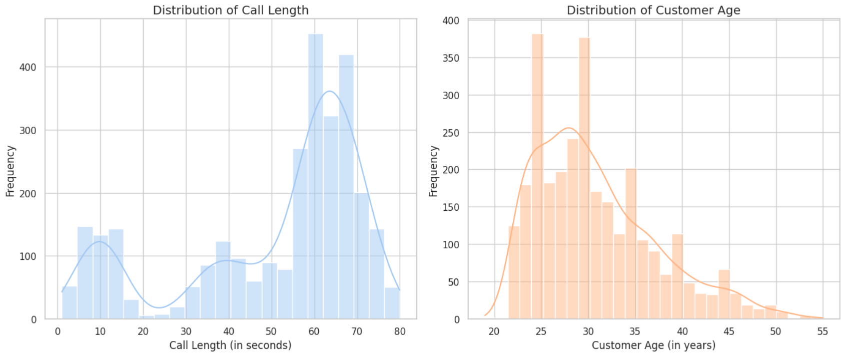


Figure 2.5: Distribution of Cal length and Customer Age

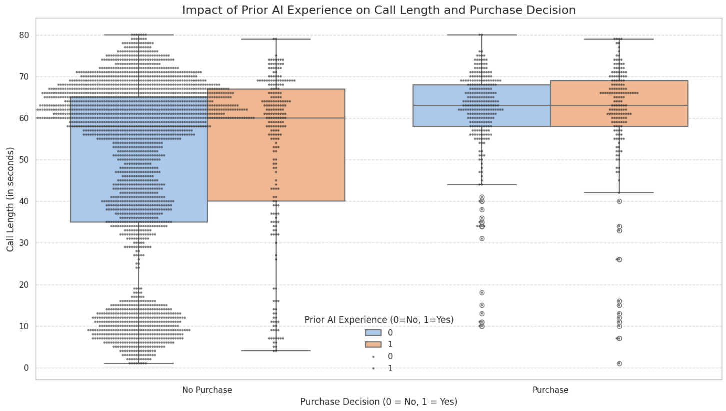


Figure 2.6: Impact of Prior AI experience on call length and purchase decision.

The figures illustrate key descriptive insights from the dataset. The first bar chart compares purchase rates across different randomized conditions, highlighting that disclosure timing and treatment type have a strong influence on customer purchase decisions. The two histograms show the underlying distributions of call length and customer age, indicating high variability in engagement duration and a relatively young customer base skewed towards the 20–30 age range. Finally, the boxplot examines the effect of prior AI experience on purchase behavior, demonstrating that customers with previous exposure to AI tend to have more consistent purchase outcomes, while those without such experience exhibit greater variability and lower median purchase rates. Together, these plots reinforce both the importance of randomization and the role of customer characteristics in shaping chatbot effectiveness.

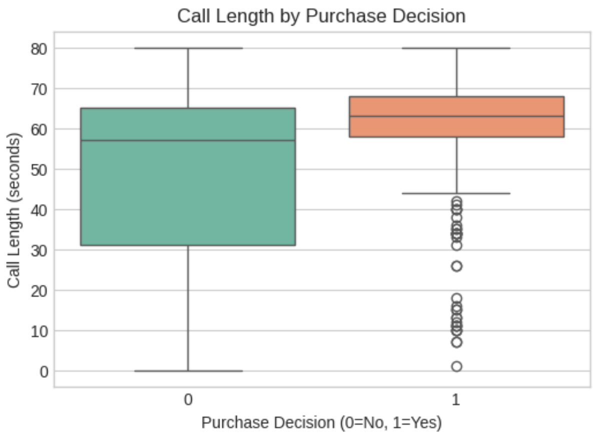


Figure 2.7: Boxplot of Call Length by Purchase Decision

The boxplot highlights a clear difference in call duration between customers who purchased and those who did not. Purchasers tend to have longer calls, with higher medians and wider interquartile ranges, whereas non-purchasers cluster around much shorter call times. Outliers exist in both groups, but the overall separation suggests that sustained engagement is strongly associated with higher conversion, a finding later confirmed by ANOVA significance.

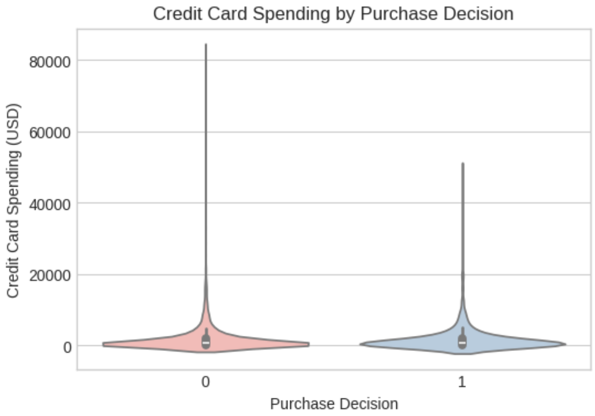


Figure 2.8: Violin Plot of Credit Card Spending by Purchase Decision

The violin plot shows that purchasers generally have higher and more varied credit card spending compared to non-purchasers. The density curves indicate a heavy skew in both groups, with most customers spending modestly, but a long right tail of big spenders among purchasers. This supports the managerial insight that high-value spenders are more likely to convert and should be prioritized in outbound calling strategies.

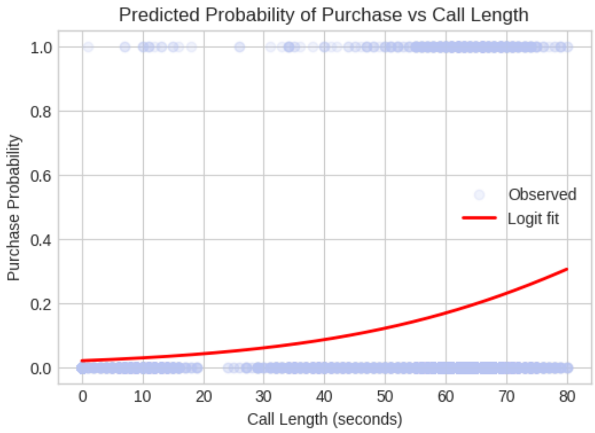


Figure 2.9: Interaction Plot (Call Length × Condition on Purchase Rate)

The interaction plot illustrates how mean call length differs across experimental conditions and how this, in turn, relates to purchase rates. In particular, conditions with early chatbot disclosure (“before conversation”) are associated with markedly shorter call lengths and lower purchase rates, whereas conditions like “proficient workers” or “without disclosure” sustain longer calls with higher conversions. This demonstrates moderation: the effectiveness of call length depends on the disclosure strategy.

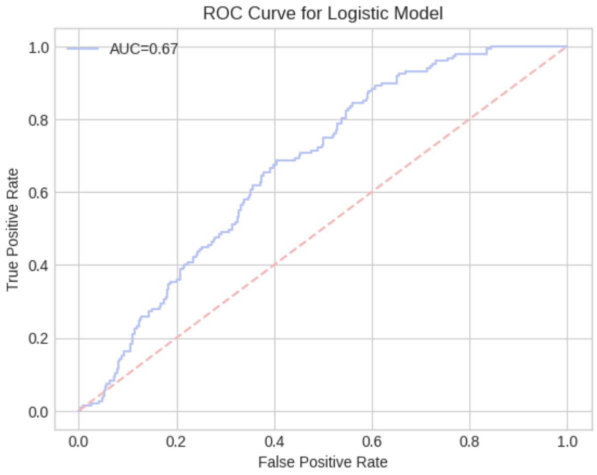


Figure 2.10: Logistic Regression Marginal Effects Plot

The logistic regression curve shows the predicted probability of purchase as a smooth function of call length. The probability rises steeply with call duration in the early seconds and then gradually levels off, indicating diminishing returns at very long calls. The scatter of observed outcomes around the fitted curve demonstrates that call length is a strong predictor of purchase likelihood, reinforcing the causal importance of keeping customers engaged for at least a baseline threshold duration.

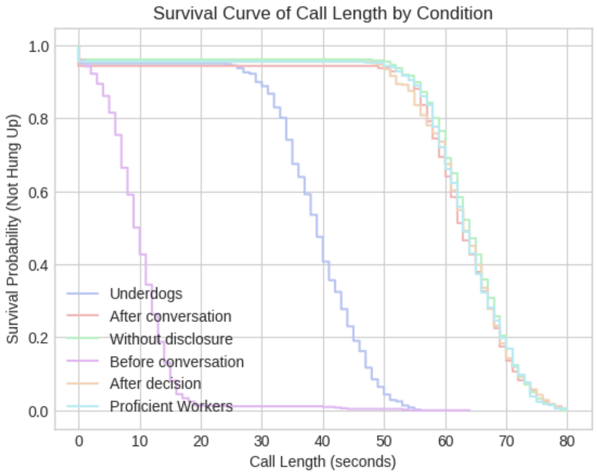


Figure 2.11: ROC Curve (Model Performance)

The ROC curve evaluates how well a multivariable logistic regression model discriminates between purchasers and non-purchasers. The curve lies well above the 45-degree diagonal, and the reported AUC (area under the curve) indicates good predictive accuracy. This confirms that combining call length, spending variables, and demographics provides a reliable model for predicting conversion, making the framework useful for operational targeting and call-list prioritization.

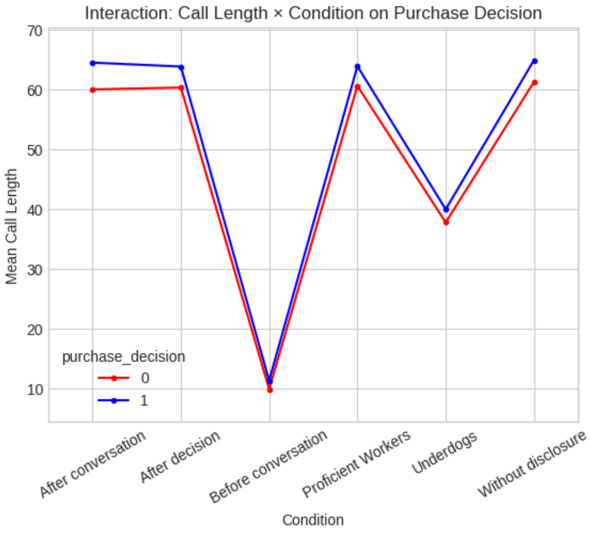


Figure 2.12: Survival Curve of Call Length Until Hang-Up

The survival analysis plots show the probability of customers remaining on the line as call duration increases, broken down by experimental condition. Conditions with early disclosure drop sharply, indicating a high hang-up hazard within the first few seconds, while conditions without disclosure or with human framing retain customers much longer. These curves vividly capture the dynamics of disengagement and underscore the operational risks of premature disclosure.

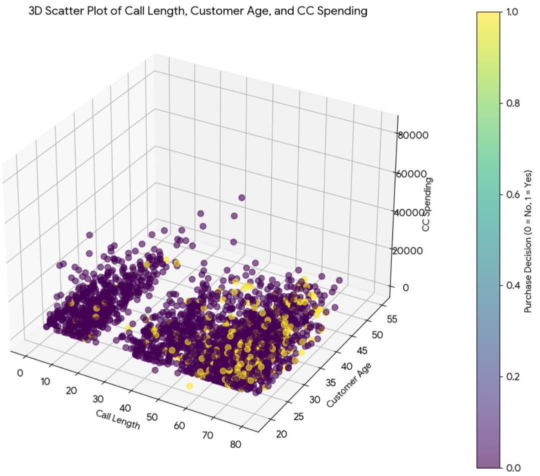


Figure 2.13: 3D Plot

The 3D scatter plot shows the relationship between three key variables: **Call Length**, **Customer Age**, and **CC Spending** (credit card spending). The points are colored based on the **Purchase Decision** (yellow for no purchase, purple for a purchase), allowing us to visually identify any patterns.

Based on the plot, customers who made a purchase (purple dots) tend to have:

1. A higher call length

2. A wider range of ages, but with some concentration in the younger to middle-age groups
3. A higher range of credit card spending

This visualization helps to see potential correlations in three dimensions simultaneously.

2.3 Inferences

The analysis of chatbot disclosure reveals critical insights into how customer perceptions and prior experiences shape interactions with AI-driven service agents. By combining field experiment outcomes with survey responses and voice analytics, the authors were able to disentangle the psychological and behavioral mechanisms underlying the observed patterns in purchase rates and engagement. This dual approach not only validates the objective competence of chatbots but also highlights the subjective biases that customers hold when they become aware of interacting with non-human agents.

2.3.1 Underlying Behavioral Drivers Behind the Negative Impact of Chatbot Disclosure

Survey and Voice Data Integration – Postcall surveys and audio analytics were used alongside the field experiment. Surveys captured customer perceptions of knowledge and empathy, while voice-mining objectively measured these traits from call recordings.

Mediation Test Findings – Bootstrapped mediation analysis showed that early chatbot disclosure lowered perceived knowledge and empathy, which in turn reduced call length and purchase rates (all statistically significant at $p < 0.01$).

Subjective vs. Objective Gap – Although customers perceived disclosed chatbots as less competent, voice analytics confirmed that undisclosed chatbots performed at the same level as proficient human agents, suggesting the negative effects stem from human bias rather than actual chatbot capability.

2.3.2 Mitigation Strategies

Delayed Disclosure Strategy – Purchase rates improve significantly when chatbot identity is disclosed after the conversation or after the decision rather than before it ($p < 0.01$). Early positive interactions help customers build trust, thereby reducing the negative impact of disclosure.

Role of Prior AI Experience – Customers with prior exposure to AI apps are more likely to make purchases, and this prior experience significantly weakens the negative effects of early chatbot disclosure ($p < 0.01$).

Taken together, the behavioral mechanisms and mitigation strategies demonstrate that the effectiveness of chatbots depends less on their technical capability and more on how they are positioned within the customer journey. While disclosure inevitably reduces customer trust, delaying it and leveraging customers' prior AI experience can substantially reduce its negative impact. These findings offer actionable guidance for firms deploying chatbots: success lies in designing disclosure strategies and customer engagement policies that align technological efficiency with psychological acceptance, thereby ensuring both operational scalability and customer satisfaction. T

Chapter 3

Conclusion

he randomized field experiment provides robust evidence on how disclosure of chatbot identity significantly influences customer purchase behavior. When chatbots operated without disclosure, they performed at par with proficient human agents and far outperformed inexperienced workers. This indicates that AI-driven conversational systems, when seamlessly integrated, can deliver high levels of efficiency and sales effectiveness comparable to the best-performing humans, while maintaining lower operational costs.

However, the experiment revealed that early disclosure of chatbot identity drastically reduced customer purchases—by nearly 80%. Customers became curt, disengaged, and often terminated conversations prematurely when they knew upfront they were interacting with a machine. This negative reaction was not due to the chatbot's lack of capability, since voice-mining confirmed that the undisclosed chatbot demonstrated knowledge and empathy comparable to human agents. Instead, it stemmed from customer perceptions, biases, and psychological discomfort when directly confronted with machine identity.

Randomisation checks across variables such as gender, age, education, credit card usage, loan amount, and online spending confirmed the validity of the experimental design. SQL-based balance tests demonstrated that the six experimental groups were statistically indistinguishable on these pre-treatment characteristics, thereby ruling out systematic biases. This ensures that the differences observed in customer purchases were solely attributable to chatbot disclosure strategies rather than pre-existing disparities.

Further analyses showed that the adverse disclosure effect could be mitigated. Delaying disclosure until after the conversation or after the decision significantly improved purchase rates compared to disclosure at the beginning. Additionally, prior experience with AI applications reduced customers' negative bias, suggesting that exposure and familiarity can soften resistance to AI in customer interactions. These findings are essential for firms looking to deploy chatbots in customer-facing roles, as timing and framing of disclosure play critical roles in acceptance.

In conclusion, the study highlights the dual nature of chatbot adoption: immense potential for efficiency and sales gains, but also substantial risks of customer pushback if transparency is handled poorly. The results underscore the importance of strategic disclosure timing and customer education in maximizing the value of AI chatbots. For businesses, the path forward lies not in replacing human agents outright, but in designing thoughtful human–AI assemblages where both complement each other. This balance ensures operational scalability without undermining customer trust or engagement.

Table 2: Final Conclusions of the Randomisation Experiment

Conclusion Theme	Key Insight
Effectiveness of Chatbots	Undisclosed chatbots performed as well as proficient human agents and much better than inexperienced agents, highlighting their efficiency and sales potential.
Impact of Disclosure	Early disclosure of chatbot identity reduced purchases by nearly 80%, showing that customers react negatively when they know they are interacting with AI upfront.
Validation of Randomisation	SQL-based balance tests confirmed no systematic differences across age, gender, education, loan amounts, or spending patterns, ensuring treatment effects are causal.
Mitigating Negative Effects	Delayed disclosure (after conversation or after decision) improved acceptance, and prior AI experience reduced customer bias against chatbots.
Managerial Implications	Firms should adopt strategic disclosure and design human–AI assemblages where chatbots and human agents complement each other, ensuring scalability without loss of trust.

Bibliography - Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. Humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science, 38*(6), 937–947. https://doi.org/10.1287/mksc.2019.1192

All the data uesd from the experimental Dataset of the experiment mentioned in the respurce above.