OPM-6600-01: Pricing and Revenue Analytics

Case Nomis A and B

Jinal Patel

Tier	Date Approved	Term (Mos.)	Amount ($)	Car Type	Comp. APR(aka Competition Rate)	Prime Rate (aka Cost of Funds)	Partner Bin	FICO
1	11/19/2004	60	18,000	U	4.85	2.13	1	705
2	11/20/2004	60	25,000	U	4.85	2.13	1	705

Once you download the Nomis case excel file, you will find multiple columns of variables. The most important column I (i.e., dependent variable) indicates whether a customer has accepted the loan or not. Other columns contain key information about the customer and the loan (i.e., independent variables). **Your first task should be identifying the relationship between the loan acceptance and other variables.** Some variables may have significant impact on the loan acceptance, while others' impact may be limited. It is common to use a regression model to analyze this relationship. Given that the dependent variable (i.e., the loan acceptance) is a binary variable (i.e., 1 = accept and 0 = otherwise).

Summary statistics (Quantitative data):

	Observations	Obs. with missing data	Minimum	Maximum	Mean	Std. deviation
FICO	1540	0	684	712	698.1084	8.351788
Amount	1540	0	17808	25000	21737.1	2292.217
Rate	1540	0	4.08	12.33	6.449357	1.297707
Term	1540	0	60	60	60	0

Summary statistics (Qualitative data):

Variable	Categories	Counts	Frequencies	%
Outcome	0	996	996	64.67532
Outcome	1	544	544	35.32468
Partner Bin	1 (Direct auto finance)	666	666	43.24675
Partner Bin	2 (Partner A)	331	331	21.49351
Partner Bin	3 (Other Partners)	543	543	35.25974
Tier	3	787	787	51.1039
Tier	2	680	680	44.15584
Tier	1	70	70	4.545455
Tier	4	3	3	0.194805

Correlation Matrix:

```
> # Print the correlation matrix
> print(correlation_matrix)
                         Tier        FICO Term       Amount Competition.Rate     Outcome        Rate Cost.of.Funds  Partner.Bin
Tier              1.000000000 -0.63999261   NA -0.001302448       0.02211934 -0.08895558  0.54823268    0.04748453  0.047366445
FICO             -0.639992605  1.00000000   NA -0.011435670       0.01509146  0.06883048 -0.41696533   -0.01468414 -0.027380884
Term                       NA          NA   NA           NA               NA          NA          NA            NA           NA
Amount           -0.001302448 -0.01143567   NA  1.000000000       0.01014534 -0.20803894  0.12837056    0.02046637 -0.005963174
Competition.Rate  0.022119337  0.01509146   NA  0.010145335       1.00000000 -0.04361446  0.15294678    0.42244396  0.012459333
Outcome          -0.088955583  0.06883048   NA -0.208038936      -0.04361446  1.00000000 -0.48065028   -0.19132659 -0.062429175
Rate              0.548232684 -0.41696533   NA  0.128370556       0.15294678 -0.48065028  1.00000000    0.38911840  0.046258482
Cost.of.Funds     0.047484528 -0.01468414   NA  0.020466365       0.42244396 -0.19132659  0.38911840    1.00000000  0.039858067
Partner.Bin       0.047366445 -0.02738088   NA -0.005963174       0.01245933 -0.06242917  0.04625848    0.03985807  1.000000000
```

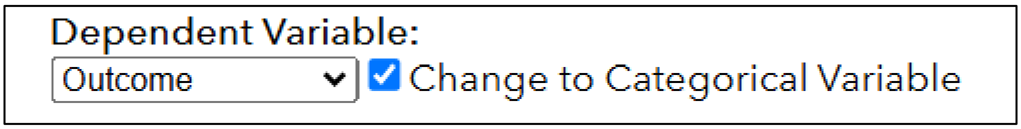
The correlation coefficient ranges from -1 to +1:

- +1 indicates a perfect positive correlation: as one variable increases, the other also increases proportionally.
- -1 indicates a perfect negative correlation: as one variable increases, the other decreases proportionally.
- 0 means no correlation: changes in one variable are not associated with changes in the other.

Logistic Regression Using R Language:

In this logistic regression, we have chosen the following:

- **Size of Training Data:** 1540 rows.
- **Dependent Variable:** Outcome.



- **In Dependent Variable:** Rate, FICO Score, Amount, Partner Bin 2 <u>(These are decided after many iterations and the best result also called Champion result and code is shown next)</u>
- We have not considered the following variables in the extracted data from the start: Term, Car Type, Approved Date, Previous Rate.
- Partner Bin variable: we separated Partner.Bin into binary columns in Excel as logistic regression categorical variables are typically transformed into binary indicators if they have more than two levels.

R Code is as follows:

```
dataset <- read.csv("Nomis.csv")
nrow(dataset)
ncol(dataset)
head(dataset)
trainingsize <- as.integer((1546-6)*(1-0/100))
validationsize <- (1546-6)-trainingsize
testingsize <- nrow(dataset)-(1546-6)
training <- head(dataset,trainingsize)
validation <- tail(head(dataset,trainingsize+validationsize),validationsize)
testing <- tail(dataset,nrow(dataset)-(trainingsize+validationsize))
if(validationsize==0) {validation <- training}
nrow(training)
nrow(validation)
nrow(testing)
model <- glm(Outcome~FICO+Amount+Rate+Partner.Bin.2,data=training,family="binomial")
summary(model)
validation$predictedLR <- predict(model,validation,type="response")
predictedclass <- ifelse(validation$predictedLR > 0.5,1,0)
confusionmatrix <- table(factor(predictedclass,levels=0:1),factor(validation$Outcome,levels=0:1))
confusionmatrix
accuracy <- (1 - mean(predictedclass != validation$Outcome,na.rm=TRUE))*100
paste(round(accuracy,2),"%",sep="")
validation$predictedLR <- NULL
predictedLR <- predict(model,rbind(validation,testing),type="response")
predictedLR <- ifelse(predictedLR > 0.5,1,0)
testing <- rbind(validation,testing)
testing <- cbind(testing,predictedLR)
write.csv(testing,"Nomis_Prediction.csv",row.names=FALSE)

#Calculate the predicted probabilities for all rows in the dataset
dataset$predicted_probability <- predict(model, dataset, type = "response")

# Save the modified dataset with the new 'predicted_probability' column to a CSV file
write.csv(dataset, "Nomis_with_Probabilities.csv", row.names = FALSE)
```

Result:

```
> dataset <- read.csv("Nomis.csv")
> nrow(dataset)
[1] 1546
> ncol(dataset)
[1] 12
> head(dataset)
  Tier FICO Term Amount Competition.Rate Outcome Rate Cost.of.Funds Partner.Bin Parner.Bin.1 Partner.Bin.2 Partner.Bin.3
1    2  702   60  22000             5.85       0 6.19          1.84           3            0             0             1
2    2  710   60  21000             5.85       0 6.19          1.84           1            1             0             0
3    3  693   60  19598             5.85       1 7.29          1.84           1            1             0             0
4    3  696   60  23071             5.85       0 7.29          1.84           3            0             0             1
5    3  697   60  21578             5.80       1 7.29          1.84           2            0             1             0
6    2  702   60  20211             5.80       1 6.19          1.84           2            0             1             0
> trainingsize <- as.integer((1546-6)*(1-0/100))
> validationsize <- (1546-6)-trainingsize
> testingsize <- nrow(dataset)-(1546-6)
> training <- head(dataset,trainingsize)
> validation <- tail(head(dataset,trainingsize+validationsize),validationsize)
> testing <- tail(dataset,nrow(dataset)-(trainingsize+validationsize))
> if(validationsize==0) {validation <- training}
> nrow(training)
[1] 1540
> nrow(validation)
[1] 1540
> nrow(testing)
[1] 6
> model <- glm(Outcome~FICO+Amount+Rate+Partner.Bin.2,data=training,family="binomial")
> summary(model)

Call:
glm(formula = Outcome ~ FICO + Amount + Rate + Partner.Bin.2,
    family = "binomial", data = training)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.831e+01  5.904e+00   6.489 8.63e-11 ***
FICO          -3.951e-02  8.177e-03  -4.831 1.36e-06 ***
Amount        -1.777e-04  2.891e-05  -6.149 7.81e-10 ***
Rate          -1.189e+00  7.345e-02 -16.182  < 2e-16 ***
Partner.Bin.2 -7.034e-01  1.707e-01  -4.120 3.79e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 2000.3  on 1539  degrees of freedom
Residual deviance: 1482.1  on 1535  degrees of freedom
AIC: 1492.1

Number of Fisher Scoring iterations: 5

> validation$predictedLR <- predict(model,validation,type="response")
> predictedclass <- ifelse(validation$predictedLR > 0.5,1,0)
> confusionmatrix <- table(factor(predictedclass,levels=0:1),factor(validation$Outcome,levels=0:1))
> confusionmatrix

      0   1
  0 863 194
  1 133 350
> accuracy <- (1 - mean(predictedclass != validation$Outcome,na.rm=TRUE))*100
> paste(round(accuracy,2),"%",sep="")
[1] "78.77%"
> validation$predictedLR <- NULL
> predictedLR <- predict(model,rbind(validation,testing),type="response")
> predictedLR <- ifelse(predictedLR > 0.5,1,0)
> testing <- rbind(validation,testing)
> testing <- cbind(testing,predictedLR)
> write.csv(testing,"Nomis_Prediction.csv",row.names=FALSE)
>
> #Calculate the predicted probabilities for all rows in the dataset
> dataset$predicted_probability <- predict(model, dataset, type = "response")
>
> # Save the modified dataset with the new 'predicted_probability' column to a CSV file
> write.csv(dataset, "Nomis_with_Probabilities.csv", row.names = FALSE)
```

Accuracy: 78.77%

- **Intercept:** The estimated intercept is 38.13, which is the baseline log-odds of the outcome when all predictors (FICO, Amount, Rate, and Partner.Bin.2) are zero. This high value suggests that without accounting for these predictors, there's an inherent bias in favor of the outcome.
- **FICO Score:** The coefficient for FICO is −0.03951, with a significant p-value, indicating that FICO has a statistically significant negative relationship with the outcome. For each unit increase in FICO, the log-odds of the outcome decrease by 0.03951. This implies that a higher FICO score reduces the likelihood of the outcome.
- **Amount:** The coefficient for Amount is −0.0001777, also statistically significant. As the loan amount increases, the probability of the outcome decreases slightly.
- **Rate:** The coefficient for Rate is −1.189 with an extremely significant p-value. This negative coefficient means that a higher rate (APR) significantly reduces the likelihood of acceptance (outcome). Each unit increase in Rate decreases the log-odds by 1.189.
- **Partner.Bin.2:** The coefficient for Partner.Bin.2 is −0.7034, also statistically significant indicating that being in this category has a negative effect on the likelihood of the outcome.

Summary:

- The predictors (FICO, Amount, Rate, and Partner.Bin.2) are all statistically significant, suggesting they have a meaningful impact on the outcome.
- Rate has the largest negative effect, indicating that higher APRs significantly reduce the likelihood of the outcome.
- The model achieves an accuracy of 78.77%, which is reasonably high, but there are still misclassifications (false positives and false negatives), which could be further analyzed or improved.

<u>Excel sheet shows the predicted value for outcome and probability.</u>

For a logistic regression model, the probability P(Y=1) is given by:

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}$$

- $\beta 0$ is the intercept,
- $\beta 1$, $\beta 2$,…, βn are the coefficients for each predictor X1,X2,…,Xn
- e is the base of the natural logarithm.

```r
# Calculate the predicted probabilities for all rows in the dataset
dataset$predicted_probability <- predict(model, dataset, type = "response")

# Save the modified dataset with the new 'predicted_probability' column to a CSV file
write.csv(dataset, "Nomis_with_Probabilities.csv", row.names = FALSE)
```

1. **What is the probability that each of these loans applicants will convert if they are quoted an APR (aka Rate) of 6.00%?**

   For **Q1 and Q2** the data is also shown in the excel sheet and pasted below.

Tier	Amount ($)	Comp. APR	Prime Rate	Partner Bin	FICO	Rate (APR)	Predicted Outcome	Predicted Probability (P=1)
1	18,000	4.85	2.13	1	705	**6**	1	0.531098175
2	25,000	4.85	2.13	1	705	**6**	0	0.246076712

- Applicant 1 has a probability of conversion (acceptance of the loan) of approximately 53.10% when quoted an APR of 6.00%.
- Applicant 2 has a probability of conversion of approximately 24.60% when quoted an APR of 6.00%.

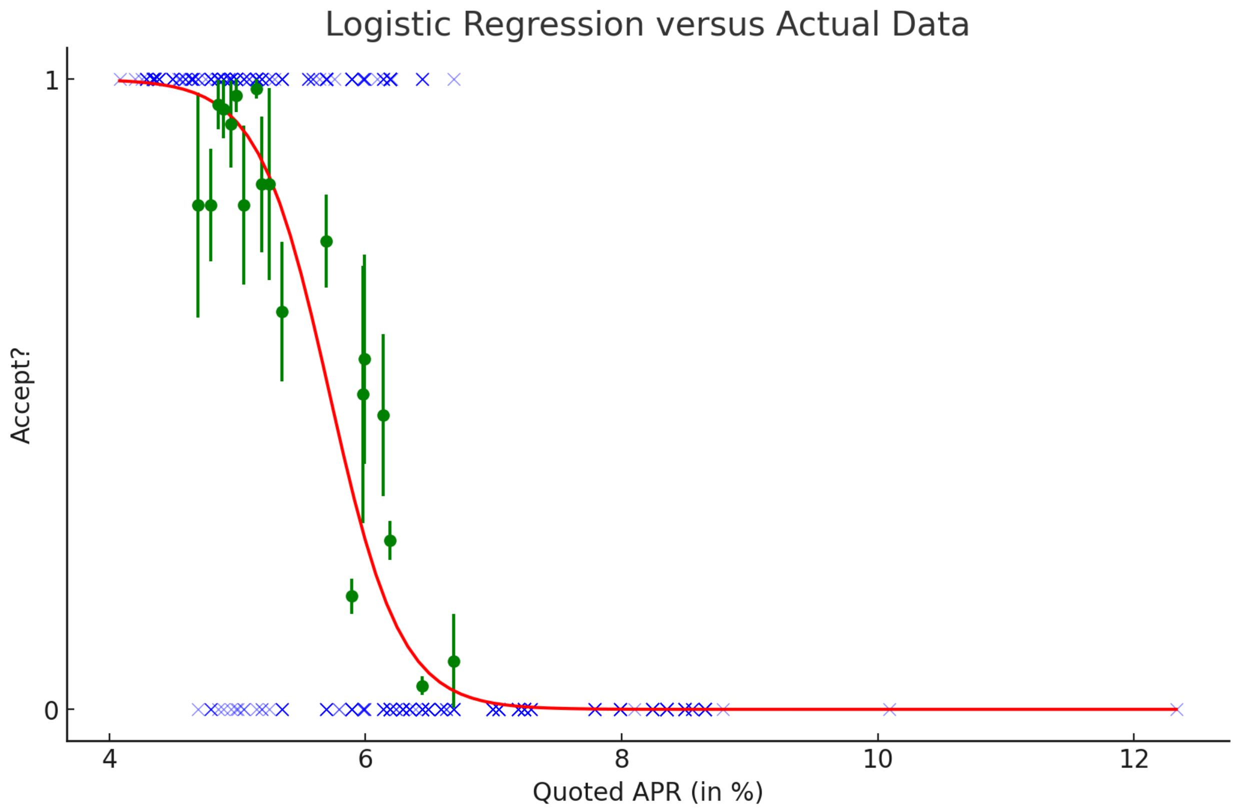
2. **How would this probability change if we changed the APRs (aka Rates) to 5.00%? To 7.00%?**

   For **Q1 and Q2** the data is also shown in the excel sheet and pasted below.

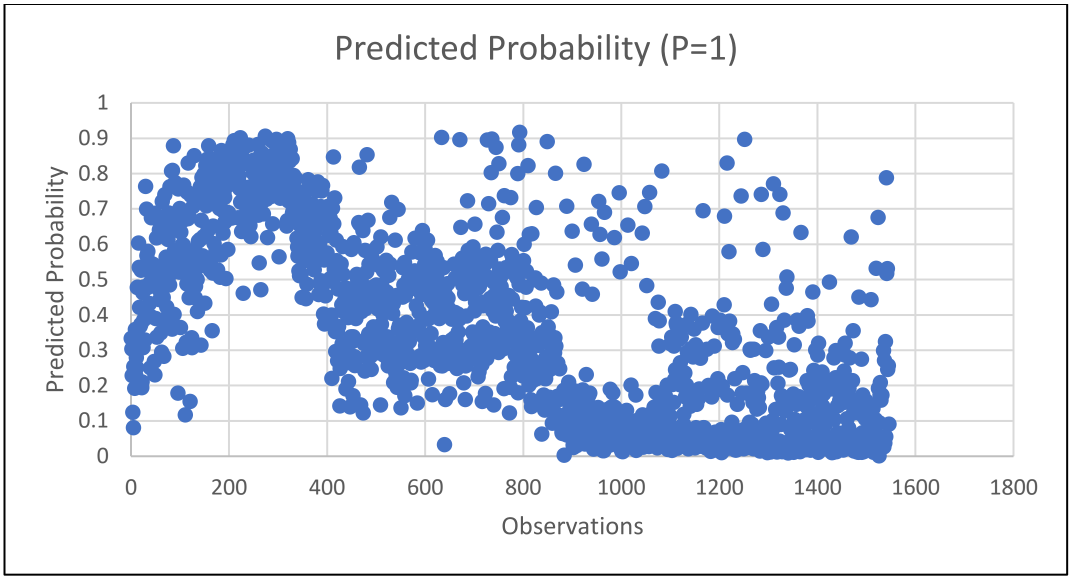
Tier	Amount ($)	Comp. APR	Prime Rate	Partner Bin	FICO	Rate (APR)	Predicted Outcome	Predicted Probability (P=1)
1	18,000	4.85	2.13	1	705	**5**	1	0.788047269
2	25,000	4.85	2.13	1	705	**5**	1	0.517241666
1	18,000	4.85	2.13	1	705	**7**	0	0.256529037
2	25,000	4.85	2.13	1	705	**7**	0	0.090438824

- Applicant 1 has a probability of conversion (acceptance of the loan) of approximately 78.80% when quoted an APR of 5.00%.
- Applicant 2 has a probability of conversion of approximately 51.72% when quoted an APR of 5.00%.
- Applicant 1 has a probability of conversion (acceptance of the loan) of approximately 25.65% when quoted an APR of 7.00%.
- Applicant 2 has a probability of conversion of approximately 9.04% when quoted an APR of 7.00%.

Higher APR makes the loan offer less attractive, leading to a lower likelihood of acceptance for both applicants.

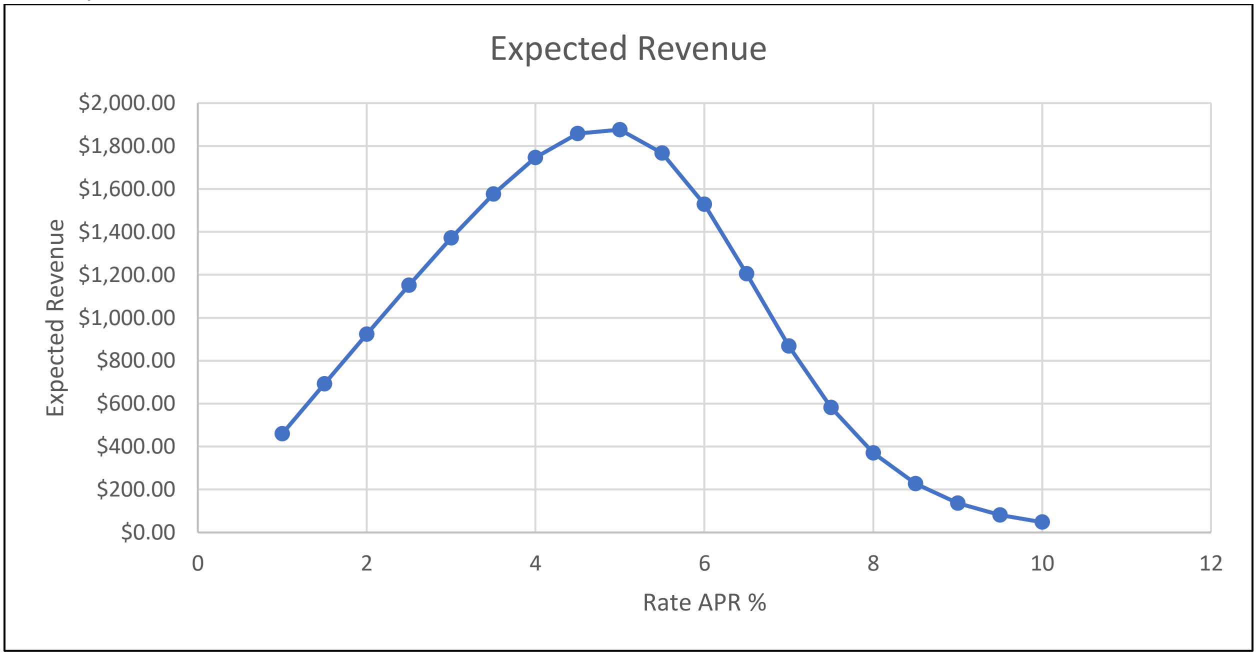


The plot shows a clear negative relationship between APR and loan acceptance probability, visually confirming the results from Questions 1 and 2. In Question 1, we calculated specific acceptance probabilities for given APRs (e.g., 6.00%). The plot supports this, as the logistic curve indicates a moderately high acceptance probability around a 6% APR. In Question 2, we examined how changes in APR (e.g., increasing to 7% or decreasing to 5%) affect acceptance. The steep decline in the logistic curve illustrates how higher APRs significantly reduce acceptance probabilities, while lower APRs increase them. The green mean points and error bars further validate this trend by summarizing acceptance rates in different APR ranges, confirming the sensitivity observed in our calculations.



3. **How should e-Car set the APRs (aka Rates) for these applications to maximize expected profit?**

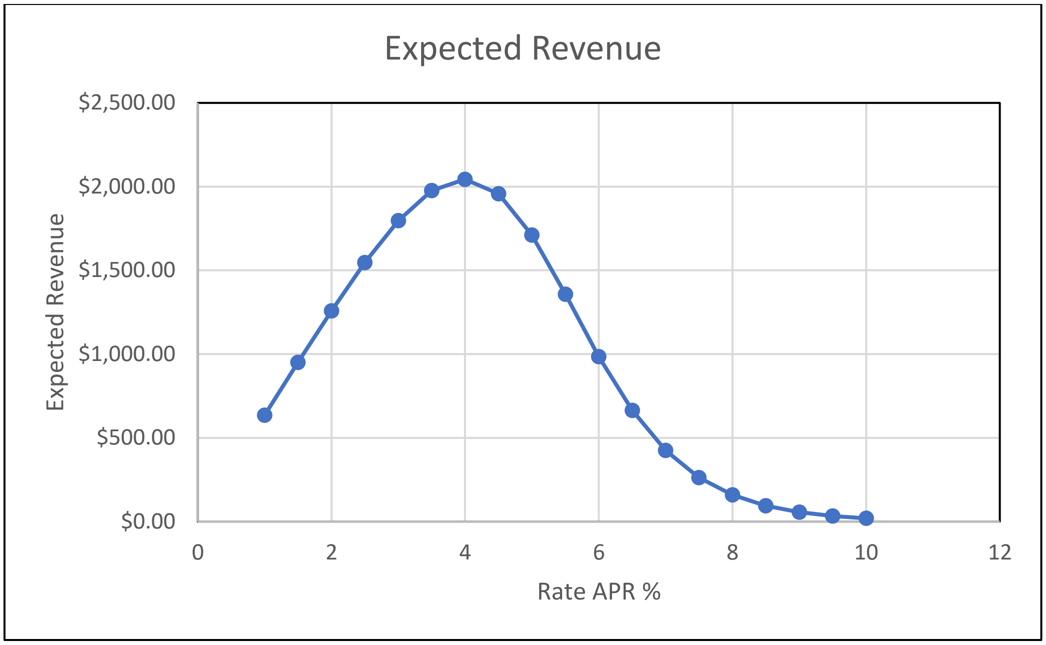
For 18,000 Loan Amounts:



	A	B	C	D	E	F	G
1	Apr(rate)	Probability	Total Interest	Total Payments	Expected Revenue		
10	5	0.78804727	$2,380.93	$20,380.93	$1,876.29	*Optimal APR= 5%*	

The optimal APR is 5%. E-Car should set the APR to 5% to maximize their expected profit. Expected Revenue: **$1,876.29**

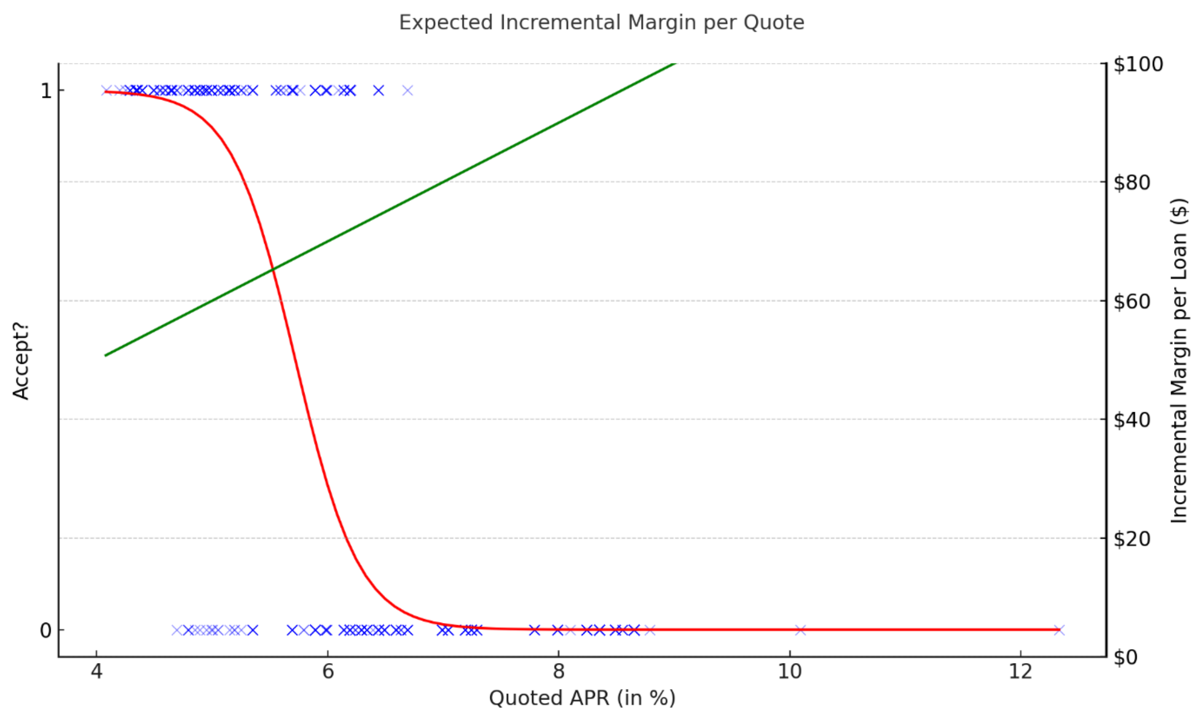
For 25,000 Loan Amount:



	A	B	C	D	E	F
1	Apr(rate)	Probability	Total Interest	Total Payments	Expected Revenue	
8	4	0.778618776	$2,624.78	$27,624.78	$2,043.71	*Optimal APR= 4%*

The optimal APR is 4%. E-Car should set the APR to 4% to maximize their expected profit. Expected Revenue: **$2,043.71**

4. How should e-Car set the APRs (aka Rates) for all loan applicants to maximize expected profit?



The plot demonstrates the relationship between APR (Quoted APR in %) and loan acceptance probability, as well as the expected incremental margin per loan. The red logistic curve indicates the likelihood of customer acceptance at different APR levels, while the green line represents the potential incremental margin per loan as APR increases.

The plot also shows that as APR increases, the probability of loan acceptance decreases, as evidenced by the downward trend of the red logistic curve. This highlights the importance of balancing APRs to maximize expected profit by attracting more customers (high acceptance rates) while also maintaining profitability (incremental margin per loan, as shown by the green line). Lowering APRs may improve acceptance rates, but this must be weighed against profitability, as indicated by the green line's increasing slope. To achieve this, we can focus on:

- **Segment-Based Pricing:** Need to use customer segments like FICO score tiers, loan amount, or partner bins to set APRs. Customers with higher FICO scores or larger loans, for example, may be more sensitive to APR changes. Offering slightly lower APRs to these segments can increase acceptance rates, improving overall profitability.
- **Competitive Benchmarking:** Since the competition rate (Comp. APR) is provided, e-Car should aim to keep APRs close to or slightly below these rates for competitive advantage. Maintaining APRs just below competitors can attract price-sensitive customers, while those less sensitive might accept higher rates.
- **Cost of Funds Consideration:** The prime rate, or cost of funds, provides a baseline below which rates cannot be profitably set. Rates should be set to ensure a positive margin above this rate, balancing profitability with the acceptance rate to avoid setting rates too close to cost, which can lead to low returns.
- **APR Elasticity:** Since the model shows APR significantly impacts acceptance probability, APRs should be set lower for customers with a high probability of conversion (those with favorable credit profiles). This could increase conversions, especially in high-value loans.
- **Dynamic and Customized Pricing:** Consider dynamically adjusting APRs based on market conditions or competitive changes. For example, if competitor rates decrease, dynamically adjusting APRs downward for a subset of price-sensitive customers could maintain e-Car’s market share.

Your second task is to quantify the impact of APR (i.e., the rate column) on the profit. It is easy to see that a higher APR generally leads a lower loan acceptance probability but a higher loan profit if accepted. I suggest your reading the book chapter of Customized Pricing by Dr. Phillips (digital copy free via SLU library)**.** Equation 13.1 on page 343 is especially crucial. This book chapter also contains many other interesting and important concepts such as endogeneity and customized pricing in action.

Expected contribution at APR p = (contribution at p) × (probability of loan acceptance at p)

Where:

- **Contribution at APR p:** This is the profit per loan in the event the applicant accepts the offer made to him/her.

- **Probability for Loan Acceptance at APR p:** This is the probability that an applicant will accept the loan at that given APR.

Contribution of different APRs: Contribution = (APR − Cost of Funds) × Loan Amount

Where:
- **APR**: The interest rate offered to the customer can be adjusted in the decimal format.
- **Cost of Funds**: The prime rate can be represented often by the lender's cost to finance the loan.
- **Loan Amount**: The principal amount of the loan.

Expected Contribution for Each APR: We can take APR 4% to 7% to get the expected contribution we should take the contribution by the probability of acceptance.

Conclusions and optimal APR decision: While comparing the expected contributions across APRs, to determine which APR maximizes expected profit.

<u>Loan amount: $18000:</u>
Cost of funds: (PRIME RATE): 2.13% or 0.0213
Estimated probability of Acceptance for APRs:

Apr(rate)	Probability
1	0.997689
1.5	0.99582077
2	0.99245367
2.5	0.98641081
3	0.97564775
3.5	0.95673394
4	0.92427044
4.5	0.87073976
5	0.78804727
5.5	0.67235905
6	0.53109817
6.5	0.38467135
7	0.25652904
7.5	0.15997575
8	0.09511431
8.5	0.05483402
9	0.03102725
9.5	0.01736653
10	0.00966041

Calculating contribution at each APR:
1) Contribution at 4% APR: (0.04−0.0213) × 18000 = 336.6
2) Contribution at 4.5% APR: (0.045−0.0213) × 18000 = 426.6
3) Contribution at 5% APR: (0.05−0.0213) × 18000 = 516.6

4) Contribution at 5.5% APR: (0.055−0.0213) × 18000 = 606.6
5) Contribution at 6% APR: (0.06−0.0213) × 18000 = 696.6
6) Contribution at 6.5% APR: (0.065−0.0213) × 18000 = 786.6
7) Contribution at 7% APR: (0.07 −0.0213) × 18000 = 876.6

Calculation expected contribution at each APR:

1) Expected Contribution at 4% APR: 336.6*0.92427044 = 311.1094
2) Expected Contribution at 4.5% APR: 426.6*0.87073976 = 371.4576
3) Expected Contribution at 5% APR: 516.6*0.78804727 = 407.1052
4) Expected Contribution at 5.5% APR: 606.6*0.67235905 = 407.853
5) Expected Contribution at 6% APR: 696.6*0.53109817 = 369.963
6) Expected Contribution at 6.5% APR: 786.6*0.38467135 = 302.5825
7) Expected Contribution at 7% APR: 876.6*0.25652904 = 224.8734

The expected contribution is the highest at 5.5% APR, rate will maximize profit under the current assumptions.

<u>Loan amount: $25000:</u>
Cost of funds: (PRIME RATE): 2.13% or 0.0213
Estimated probability of Acceptance for APRs:

Apr(rate)	Probability
1	0.99202597
1.5	0.985645576
2	0.974292229
2.5	0.954374846
3	0.920288744
3.5	0.864356596
4	0.778618776
4.5	0.660004592
5	0.517241666
5.5	0.371607533
6	0.246076712
6.5	0.152649761
7	0.090438824
7.5	0.052024728
8	0.02939971
8.5	0.016443407
9	0.009143094
9.5	0.005067174
10	0.002803127

Calculating contribution at each APR:

1) Contribution at 4% APR: (0.04−0.0213) × 25000 = 467.5
2) Contribution at 4.5% APR: (0.045−0.0213) × 25000 = 592.5

3) Contribution at 5% APR: (0.05−0.0213) × 25000 = 717.5
4) Contribution at 5.5% APR: (0.055−0.0213) × 25000 = 842.5
5) Contribution at 6% APR: (0.06−0.0213) × 25000 = 967.5
6) Contribution at 6.5% APR: (0.065−0.0213) × 25000 = 1092.5
7) Contribution at 7% APR: (0.07 −0.0213) × 25000 = 1217.5

Calculation expected contribution at each APR:

1) Expected Contribution at 4% APR: 467.5*077862 = 364.0043
2) Expected Contribution at 4.5% APR: 592.5*0.66 = 391.0527
3) Expected Contribution at 5% APR: 717.5*0.51724 = 371.1209
4) Expected Contribution at 5.5% APR: 842.5*0.37161 = 313.0793
5) Expected Contribution at 6% APR: 967.5*0.24608 = 238.0792
6) Expected Contribution at 6.5% APR: 1092.5*0.15265 = 166.7699
7) Expected Contribution at 7% APR: 1217.5*0.09044 = 110.1093

For this loan of $25,000, the **highest expected contribution** is at **4.5% APR**, yielding $391.0527. Therefore, setting the APR at 4.5% would be the optimal choice for maximizing profit in this case.

Percentage of annual rate affects profit as it reflects the exact profitability of loan companies. It normally means that more profit is made as well because it attracts more customers, and hence more loans offered. Compared to its competition, the existence of a competitive APR may be one of the ultimate selling points for potential customers. Since borrowers tend to choose products with low and therefore adjustable annual percentage rates (APRs), this can translate to wider revenues and profits. On the other hand, if you establish your APR higher than all your competitors, you may discourage potential borrowers and possibly reduce your profit margin. Therefore, measuring and especially purposefully positioning the APR is crucial for getting the highest profit in the lending industry.