Name: **Suraj Sunil Pawar**
Unity ID - **spawar2**
Student ID- **200315997**

---

1) Single Predictor Regression Model

Predictor with the highest estimate = Open Price
Let Open Price = X

a. Probabilities
$\beta 0 = 1.544$
$\beta 1 = -0.440$
n = 1972
$p = 1 / (1 + e^{[-(\beta 0 - \beta 1 X)]})$
$p = 1 / (1 + e^{[-(1.544 - 0.440X)]})$
$\mu_y = np$
$\mu_y = 1972 / (1 + e^{[-(1.544 - 0.440X)]})$

b. Odds
$\text{Odds} = \mu_y / (1 - \mu_y) = 1972 / (e^{[-(1.544 - 0.440X)]} - 1971)$

c. Logit
$\text{Logit} = \log (\mu_y / (1 - \mu_y)) = log\,(1972) - log\,(e^{[-(1.544 - 0.440X)]})$
$= 7.5868 - log\,(e^{[-(1.544 - 0.440X)]} - 1971)$

2) All Predictor Regression Model

Predictors with highest absolute value estimates are Open Price, Close Price, Category_Category6, currency_Currency2.

Let,
Open Price = X1
Close Price = X2
Category_Category6 = X3
Currency_Currency2 = X4

a. Probabilities

$\beta 0 = 0.328$
$\beta 1 = -1.014$
$\beta 2 = 0.7898$
$\beta 3 = 0.5528$
$\beta 4 = 0.434$

$z = 0.328 - 1.014X1 + 0.7898X2 + 0.5528X3 + 0.434X4$
$p = 1 / (1 + e^{-z}) = 1 / (1 + e^{[-(0.328 - 1.014X1 + 0.7898X2 + 0.5528X3 + 0.434X4)]})$

μ_y = np
μ_y = 1972 / (1 + $e^{[-(0.328 - 1.014X1 + 0.7898X2 + 0.5528X3 + 0.434X4)]}$)

b. Odds
Odds = μ_y / (1 − μ_y) = 1972 / ($e^{[-(0.328 - 1.014X1 + 0.7898X2 + 0.5528X3 + 0.434X4)]}$ - 1971)

c. Logit
Logit = log (μ_y / (1 − μ_y)) = log (1972 / ($e^{[-(0.328 - 1.014X1 + 0.7898X2 + 0.5528X3 + 0.434X4)]}$ - 1971))

3) Odds Ratio

The predictor with the highest estimate is Open Price.

Odds Ratio
= ($e^{[-(0.328 - 1.014(X1) + 0.7898X2 + 0.5528X3 + 0.434X4)]}$ − 1971) /
($e^{[-(0.328 - 1.014(X1+1) + 0.7898X2 + 0.5528X3 + 0.434X4)]}$ − 1971)

= $e^{13.014}$
= 1.014

A linear regression model, the increase will be directly proportional to the Coeff that is 13.014

4) Reduced Logistic Regression Model

Statistically significant predictors considered for reduced model are 'sellerRating', 'endDay_Day3', 'endDay_Day2', 'Category_Category2', 'endDay_Day1', 'Category_Category7', 'currency_Currency2', 'Category_Category6', 'ClosePrice', 'OpenPrice'.

Accuracy of a model built with all predictors = 0.8035
Accuracy of reduced model = 0.8074

5) Dispersion of Model

Expected Variance of the Model = 0.2482

Observed Variance of Model = 0.2472

[Observed Variance is computed using *y_pred_reduced.var()*]

Ø = observed variance / Expected variance = 0.2472 / 0.2482 = 0.996

The Logistic Regression model is not overdispersed as Ø < 1.

If there was overdispersion in our model, the test would have returned positive and then we would have to refit our model with quasi-binomial distribution instead of binomial.