# **MAR 653 - Homework 2**

**November 12, 2023**

Group 2

* Anna Khairi
* Logan Roach
* Mark Stiles
* Oliver Gunn
* Sachi Singh

# Summary

* 1. Haven’t used dummy variables in logistic regression
  2. Not sure if the interpretation is worded correctly
  3. Need to predict retention rate

1. **Generate initial insights from the data.**
   1. The most correlated variable with retention is order frequency, doorstep and esent
2. **Split the data into test and train. Describe the logic behind the split.**
   1. I sorted by custid since it didn’t appear to be correlated to any of the dates, or cities which would bias in favor of a cohort or region, then I split the data in half into hopefully random groups
3. **Use esent, eclickrate, avgorder, ordfreq, paperless, refill, doorstep as independent**

**variables to estimate the model using train data.**

* 1. **Interpret the coefficients in laypersons terms.**

| Variable | Coefficient | Interpretation |
| --- | --- | --- |
| intercept | -3.217 | Constant value / no variable |
| esent | 0.224 | Increases odds of retention 3rd highest |
| eclickrate | 0.021 | Increases odds of retention slightly |
| avgorder | -0.003 | Decreases odds of retention slightly |
| ordfreq | 1.447 | Increases odds of retention the highest |
| paperless | -0.009 | Decreases odds of retention |
| refill | 0.322 | Increases odds of retention the 2nd highest |
| doorstep | 0.164 | Increases odds of retention |

* 1. **Predict retention, and calculate hit rate in the test data**
     1. Hit Rate: 97.1%

1. **Add favday and city to the above model.** 
   1. **Interpret the coefficients in laypersons terms.**
   2. **Predict retention**
   3. **Calculate hit rate in the test data**
2. **Tweak / Fine tune your results from steps 3 or 4. (Hint: Consider removing variables**

**or adding new transformed variables). Does the hit rate decrease or increase and why?**

1. **Provide recommendations for retail relay to improve retention**
   1. Provide a discount or coupon for frequent orders
   2. Include doorstep free in first delivery because it increases the odds of retention by 49%
   3. Send each customer 35-40 email notifications over the course of their lifetime, which is roughly 18 per month because it also increases the odds 27%
   4. Encourage users to sign up for automatic refill
2. **List any other data that would be helpful for the analysis**
   1. Paperless, refill, and doorstep were significantly higher for retained customers. It’s not clear if that’s the leading or trailing indicator but it could be due to the higher number of emails that informed them of the features.

# Analysis

## 

# Recommendations