Hi Jason,

Please find below project work for your perusal:

* [Executive summary](Summary.pdf) (summary.pdf)
* [Answers & analysis workflow](040-scripts/answers.html) (answers.html)

Also included in this document is the following:

* [Assumptions](#_Assumptions) – list of assumptions made mapped to associated stakeholder
* [Directory](#_Directory) – layout of project i.e. key files, folders, structure

Regards,

Swapnil

# Assumptions

* Marketing / Product
  + In users: missing gender information is assumed to be a legitimate value of non-declaration (can be converted to ‘other’ for reporting consistency). Values changed to ‘NA’
  + In users, orders: missing values for order\_channel and discovery\_channel (when order\_channel\_category or discover\_channel\_category is equal to ‘paid’) represent inconsistent data. Missing values converted to ‘other’
  + In orders: there appear to be a number of orders placed for products before their launch date. There may well be a good business reason for this but for the purposes of this analysis these line item orders were ignored. You can find the complete list [here](020-notes/question_1.xlsx)
  + The Shoe Lovers list (identified in 3a) is assumed to target loyalists with a demonstrated affinity (i.e. purchasing history) for the product category
* Technical / Data
  + ALL: raw timestamp data is assumed to be stored in universal time (UTC)

# Directory

010-data /

orders\_full.csv

products\_full.csv

users\_full.csv

020-notes /

[category user\_id lists x 14].csv

030-assets /

040-scripts /

answers.Rmd

answers.html

helpers.R

050-graphs /

data\_analysis.Rproj

START.pdf

Summary.pdf