**UDACITY – DATA SCIENTIST NANODEGREE**

**TERM II**

**PROJECT 6: CAPSTONE PROJECT - STARBUCKS**

# CONTENTS TABLE:

1. [Pre-requirements.](file:///C:\Users\Ibrahim\Downloads\README.docx#_PRE-REQUIREMENTS:)
2. [Project Motivation.](file:///C:\Users\Ibrahim\Downloads\README.docx#_PROJECT_MOTIVATION:)
3. [Description of files.](file:///C:\Users\Ibrahim\Downloads\README.docx#_DESCRIPTION_OF_FILES:)
4. [Findings](file:///C:\Users\Ibrahim\Downloads\README.docx#_FINDINGS:)
5. [License.](file:///C:\Users\Ibrahim\Downloads\README.docx#_LICENSE:)
6. [Future Work.](file:///C:\Users\Ibrahim\Downloads\README.docx#_FUTURE_WORK:)

# PRE-REQUIREMENTS:

For this project, there is no exceptional requirements, just the user needs to install python and run the tool.

The following python documentation will provide more information.

<https://docs.python.org/3/>

Aslo, as we used the matrix factorization, I recommend the user to read the details in the following link:

<https://towardsdatascience.com/overview-of-matrix-factorisation-techniques-using-python-8e3d118a9b39>

# PROJECT MOTIVATION:

Starbucks is one of the global companies that provide drinks for people overall the world, as they on regular basis provide their customers by different offers, in the project we used three different datasets to analyze the people behavior and what's the best offer for them.

As a data scientist geek, you need to understand the data and manipulate it based on different tools and techniques such as one hot encoding and remove the null values and identify some metrics such as removing the customer with age more than 100.

# DESCRIPTION OF FILES:

The below will help you to know about each of the available files:

3.1 Portfolio.json

* id (string) - offer id
* offer\_type (string) - type of offer ie BOGO, discount, informational
* difficulty (int) - minimum required spend to complete an offer
* reward (int) - reward given for completing an offer
* duration (int) - time for offer to be open, in days
* channels (list of strings)

3.2 Profile.json

* age (int) - age of the customer
* became\_member\_on (int) - date when customer created an app account
* gender (str) - gender of the customer (note some entries contain 'O' for other rather than M or F)
* id (str) - customer id
* income (float) - customer's income

3.3 transcript.json

* event (str) - record description (ie transaction, offer received, offer viewed, etc.)
* person (str) - customer id
* time (int) - time in hours since start of test. The data begins at time t=0
* value - (dict of strings) - either an offer id or transaction amount depending on the record

# FINDINGS:

Based on the analysis of the data, the model will support you by identify the best offer for each user based on the transcript and other metrics.

All the details can be available through my blog.

<https://development4geeks.wordpress.com/2019/12/15/starbucks-capstone-challenge/>

# LICENSE:

The raw data for the developers' survey is licensed for stack overflow under the [stack overflow privacy policy.](https://stackoverflow.com/legal/privacy-policy?_ga=2.188457640.1669360686.1572983339-326182324.1566123578)

# FUTURE WORK:

This script can be updated to be analysis for different information such as gender, income and age and the rate of offer completion, as many user receive the offers but didn't use it.