# National College of Ireland

## Project Submission Sheet – 2020/2021

| | |
|---|---|
| **Student Name:** | Aanchal Singh, Deepak Kumar Swain, Sai Prasanna Gontyala and Sweta Kumari<br>……………………………………………………………………………………………………… |
| **Student ID:** | 19221771, 19216769, 19233388 and 19240848<br>……………………………………………………………………………………………………… |
| **Programme:** | MSc. Data Analytics …………………………………………………… **Year:** 2020 ……………………… |
| **Module:** | Business Intelligence & Business Analytics<br>……………………………………………………………………………………………………… |
| **Lecturer:** | Vikas Sahni<br>........................................................................................................ |
| **Submission Due Date:** | 20-12-2020<br>……………………………………………………………………………………………………… |
| **Project Title:** | Business Analytics for Netflix – Specification<br>……………………………………………………………………………………………………… |
| **Word Count:** | 1899<br>……………………………………………………………………………………………………… |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

| | |
|---|---|
| **Signature:** | *Aanchal Singh, Deepak Kumar Swain, Sai Prasanna Gontyala and Sweta Kumari*<br>……………………………………………………………………………………………………… |
| **Date:** | 20-12-2020<br>……………………………………………………………………………………………………… |

### PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

| Office Use Only | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Business Analysis for OTT Platform- Netflix

Deepak Kumar Swain
*MSc Data Analytics*
*National College of Ireland*
Dublin, Ireland
x19216769@student.ncirl.ie

Sweta Kumari
*MSc Data Analytics*
*National College of Ireland*
Dublin, Ireland
x19240848@student.ncirl.ie

Sai Prasanna Gontyala
*MSc Data Analytics*
*National College of Ireland*
Dublin, Ireland
x19233388@student.ncirl.ie

Aanchal Singh
*MSc Data Analytics*
*National College of Ireland*
Dublin, Ireland
x19221771@student.ncirl.ie

*Abstract*— **With the immense usage of digital medium there has emerged a significant growth in the streaming platforms. The recent pandemic which locked the movement of people had led to a great economic crisis yet it boosted the finance of quite a few sections. One such sectors is the streaming media services. With the increase in the viewers, it is really essential for the OTT providers to monitor not only the sales but also the shows and genres that the customers would prefer the most based on various factors like location or age and cater their needs accordingly. Emphasizing on this need, we chose to implement an end-to-end business analysis on Netflix that gives a clear view to the provider on the sales and growth adhering to customer satisfaction.**

## I. Background and Scope

Netflix is one of the leading OTT platforms founded in 1997 by Reed Hastings and Marc Randolph, head quartered in California. Initially it delivered services only for DVD sales and rental later in 2007, it expanded its business to streaming media and retained the DVD business. It purchases the Television shows and movies from various production houses or finances from there in-house production in the name of 'Netflix Original' and caters the customer interests based on the subscriptions.

The current analysis focuses on six European countries - Sweden, Ireland, Czech Republic, Bulgaria, Belgium and England as this American company is currently focusing on expanding its business in the European countries that are known for entertainment. There are three subscriptions available for the customer to choose – Basic, Standard and Premium. The basic subscription that costs 8 euros provides a single screen to access the services with the quality of 720p lines of vertical resolution. The Standard subscription provides two screens to access the services with the quality of 1080p lines of vertical resolution. While a premium subscription provides 4 screens with a supreme quality of 4K+HDR. The price of standard and premium subscriptions is 12 euros and 16 euros respectively.

## II. Project Specification and Process Flow

The prominent factor of the analysis is to analyse the interests of the customers and work on customer satisfaction by reinforcing discounts to the targeted customers, thereby, bagging more subscriptions and eventually increasing the sales of Netflix.

The users can choose a suitable subscription on the application or the website. Whenever an individual shows interest in Netflix by signing up on the website, the information is stored in Microsoft Dynamics and later monitoring the further interests, a new user promo code for a first-time user or a seasonal promo code is provided to convert the opportunity to customer. The subscribed customer information from various locations is stored on cloud database. The master data is later processed and stored in Azure Database. All the details of genre, shows, subscriptions and transactions that helps the organization to keep a check on the inventory and sales are maintained and handled on the database. The database is then connected to Power BI. The section of customers to be focused on is derived from the reports created on Power BI based on the age group, watch duration and location. To keep intact with the competitors and bag great shows or movies, a bid chart is constantly tracked on Microsoft Dynamics where the bidding of shows or movies is done.
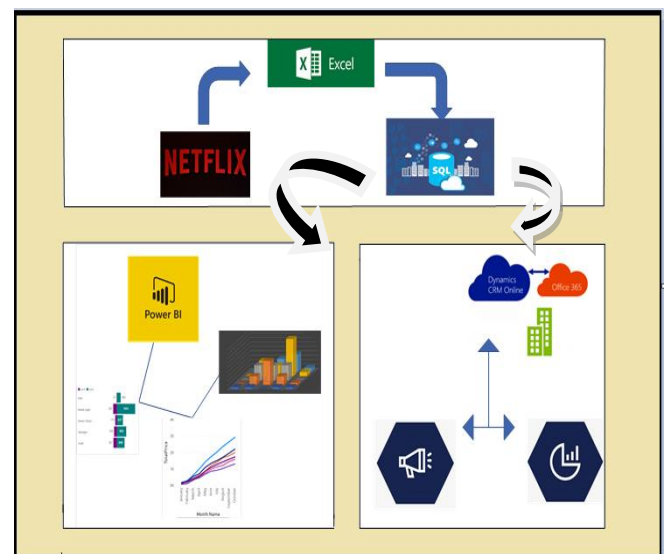


Fig. 1. Process Flow of Netflix

## III. Analytical Objectives and Requirements

Every move of the business process is monitored in the analytics as every step has its significant role in making a better business and delivering great content to the subscribers. To meet this goal of fine customer satisfaction and great sales various reports are created on Power BI. The sections on the dashboard would answer the below requirements:

- Sales in 2019 and 2020 based on each month and Quarter.

- Total subscription amount earned in each location every month.

- Subscription Analysis per every month, country and based on age group.

- Promocode usage based on each country.

- Most watched shows based on the age group.

- Top viewed movies and shows.

- Top rated movies and shows by the viewers.

- The top watched shows based on the location.

- Usage percentage of each payment mode.

- Total likes or dislikes of all the shows and movies based on the customer feedback.

## IV. DATA AND CUSTOMER INTEGRATION

The actual data of Netflix would be humungous and the data maintenance and integration would be a challenging task. For the current analysis, the master customer data is created using Mockaroo and the data related to shows, genre is taken from Kaggle [1]. While the data related to countries and currency is taken from the official website. The transactional data is created by using SQL functions. The data consists of six thousand customers and about three thousand shows. It also has data for six countries and forty-seven thousand rows of transactional data which has been created for the analysis. The master data of every subscriber would be stored on cloud on Azure database from where data is segregated to dimensions and facts.

## V. DATABASE DESIGN

A relationship type represents an association between or among several entities [2]. Entity Relationship diagram gives the clear insights of how the organization works through the relationship between the entities. Here in Netflix, we have the entities - Customer, Subscription, Shows, Genre, Promocode, Region, Currency, Customer Watch History and a transactional table, Fact Transaction.

- **Customer** – This holds the entire information of the current subscribers and the inactive subscribers like the age, gender, location, card details against the subscription date and the subscription period.

- **Subscription** – This entity contains the information like price, features of the three subscriptions – Basic, Standard, Premium that are available for the viewers.

- **Shows** – The shows table holds the details of all the shows and movies that are available on Netflix for the viewers.

- **Genre** – This table has information of all genres that are available on Netflix.

- **Promocode** – There are promocodes with some calculative discounts that are provided to the first subscriber and some seasonal offers. The details regarding this are recorded in this table.
- **Region** – The Region table holds all the data related to the countries that we have considered in the analysis.

- **Currency** – The Currency table has the details of the currencies that work in the countries on which the analysis is done.

- **Customer Watch History** – Every time a user starts watching a show or movie, an entry is made in this table that depicts the show, genre and the watch duration.

- **Transactions** – Every transaction the user makes is recorded in this table. This also holds the discount against each transaction. In case of any promocode applied on a transaction, the billing amount is calculated accordingly on which the aggregations are done to analyze the sales.
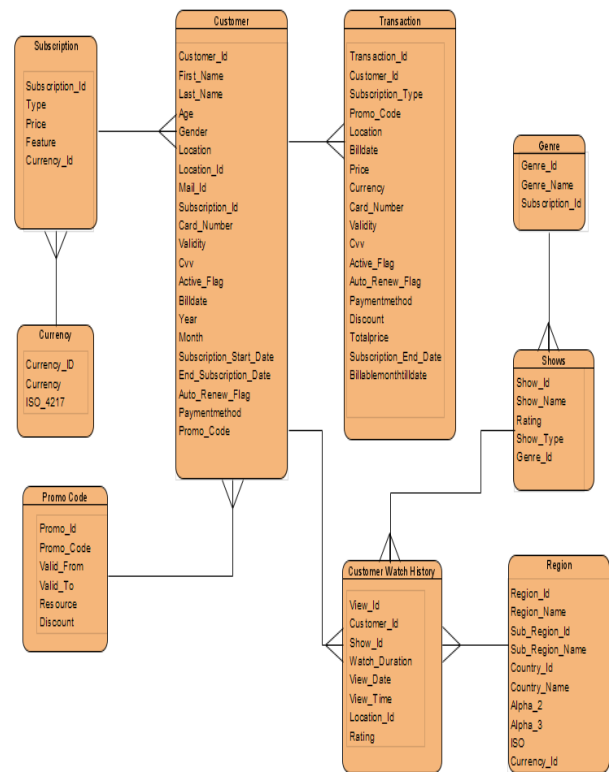


Fig. 2.  Entity Diagram of Netflix

The Customer watch history table is a major source of the watch analysis for a better delivery of desired content for every section of viewers. To meet this requirement this is connected to all the entities like shows, genre, customer from where the analysis based on age can be done. This table also has the feedback information – Rating that is given by a customer which helps in the analysis to check if the content that is being catered is ultimately liked by audience.

The transaction table where each transaction made by a subscriber is recorded paves the base for all the sales analysis. This is connected to customer that holds all the information of the subscriber which helps in analyzing the sales made in a particular year or location or the most sold shows.

The relationship between the entities is set based on the functional possibilities. The customer table holds the master data and the unique identity being the column Customer ID. As a customer can have a single subscription, but a subscription can have an ample number of customers so the relationship is chosen to be many to one. Also, a customer can have access to multiple promocodes over the period. So, the relation between the promocode to customer is chosen to be many to one. Likewise, the relationships between the entities are set based on the functional business process and would satisfy the analytical requirements.

## VI. DATA DICTIONARY

A data dictionary gives an insight at the granular level of a column on its functional significance and the columnar description. Below are the data dictionaries of each entity that is part of the analysis:

- Customer

| CUSTOMER | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| CUSTOMER_ID | int | Unique Customer ID | No |
| FIRST_NAME | varchar(50) | Customer's First Name | No |
| LAST_NAME | varchar(50) | Customer's Last Name | No |
| AGE | int | Customer's Age | No |
| GENDER | char(1) | Customer's Gender | No |
| LOCATION | varchar(100) | Customer's Residing country name | No |
| LOCATION_ID | int | Uinque Country ID | No |
| MAIL_ID | varchar(150) | Customer's email id | Yes |
| SUBSCRIPTION_ID | int | Unique Subscription ID | No |
| CARD_NUMBER | varchar(20) | Customer's Card Number | Yes |
| VALIDITY | datetime | Validity of Customer's card | No |
| CVV | int | Card Verification Value | Yes |
| ACTIVE_FLAG | char(1) | Flag to check if customer is active | No |
| BILLdate | datetime | Subscription bill date | No |
| YEAR | int | Stores the subscription year | No |
| MONTH | int | Stores the subscription month | No |
| Subscription_Start_date | datetime | Stores subscription start date | No |
| End_Subscription_date | datetime | Subscription end date. If blank, subscription is active. | No |
| AUTO_RENEW_FLAG | char(1) | If Y, auto-payment will be done from user's account. | No |
| PAYMENTMETHOD | varchar(20) | Method of payment | No |
| PROMO_CODE | varchar(10) | Promo code | No |

Fig. 3. Data Dictionary of Customer Table

- Subscription

| SUBSCRIPTION | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| SUBSCRIPTION_ID | int | Unique Subscription ID | No |
| TYPE | varchar(50) | Subscription Type | No |
| PRICE | numeric(18, 0) | Amount Per Subscription | No |
| FEATURE | varchar(50) | Features Available In Each Subscription | No |
| CURRENCY_ID | int | Unique Currency ID Taken From Currency Table As Foreign Key | No |

Fig. 4. Data Dictionary of Subscription Table

- Shows

| SHOWS | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| SHOW_ID | int | Unique Show ID | No |
| SHOW_NAME | varchar(500) | Show Name | No |
| CERTTIFICATE | varchar(500) | Certificate Received | No |
| SHOW_TYPE | varchar(500) | Show Type: Movie or TV Show | No |
| GENRE_ID | varchar(50) | Unique Genre ID taken from Genre Table as Foreign Key | No |

Fig. 5. Data Dictionary of Shows Table

- Genre

| GENRE | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| GENRE_ID | int | Unique ID for Genre | No |
| GENRE_NAME | nvarchar(300) | Genre Name | No |

Fig. 6. Data Dictionary of Genre Table

- Region

| REGION | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| Region_ID | int | Unique Region ID | No |
| Region_Name | varchar(50) | The Region in Scope: Europe | No |
| Sub_Region_ID | int | Unique ID for Sub Region | No |
| Sub_Region_Name | varchar(50) | Four Sub-Regions in Europe; North, South, East and West | No |
| Country_ID | int | Unique ID for Country | No |
| Country_Name | varchar(50) | Country Name | No |
| Alpha_2 | varchar(50) | Two-Letter Country Codes defined in ISO | No |
| Alpha_3 | varchar(50) | Three-Letter Country Codes defined in ISO | No |
| ISO | varchar(50) | ISO Code | No |
| CURRENCY_ID | int | Unique Currency ID taken from Currency table as Foreign Key | No |

Fig. 7. Data Dictionary of Region Table

- Currency

| CURRENCY | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| Currency_ID | int | Unique Currency ID | No |
| Currency | varchar(50) | Currency Name | No |
| ISO_4217 | varchar(50) | Currency Code | No |

Fig. 8. Data Dictionary of Currency Table

- Promocode

| PROMO CODE | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| PROMO_ID | int | Unique Promo ID | No |
| PROMO_CODE | varchar(20) | Promo Code Name | No |
| VALID_FROM | date | Promo Code Valid From | No |
| VALID_TO | date | Promo Code Valid To | No |
| RESOURCE | varchar(50) | Resource from where promo code received | No |
| DISCOUNT | int | Percentae Discount on Actual Payment | No |

Fig. 9. Data Dictionary of Promocode Table

- Customer Watch History

| CUSTOMER_WATCH HISTORY | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| VIEW_ID | int | Unique view ID | No |
| CUSTOMER_ID | int | Foreign key taken from Customer table | No |
| SHOW_ID | int | Foreign key taken from Shows table | No |
| WATCH_DURATION | decimal(18, 0) | Watch duration of customer | No |
| VIEW_DATE | date | Date on which the customer viewed any shows | No |
| VIEW_TIME | time(7) | Time when the user started viewing | No |
| LOCATION_ID | int | foreign key taken from region table. Country ID | No |
| RATING | int | 1, liked by user0, disliked bu user | No |

Fig. 10. Data Dictionary of Customer Watch History Table

- Transaction

| TRANSACTION | | | |
|---|---|---|---|
| Attribute Name | Datatype | Description | Data Masking |
| TRANSACTION_ID | int | Unique Transaction ID | No |
| CUSTOMER_ID | int | Foreign key taken from Customer table | No |
| SUBSCRIPTION_type | varchar(100) | Type of Netflix subscription | No |
| PROMO_CODE | varchar(10) | Promocode names | No |
| LOCATION | varchar(200) | Country Name | No |
| BILLDATE | datetime | Subsription payment date | No |
| PRICE | numeric(18, 0) | Subscription payment amount | No |
| CURRENCY | varchar(100) | Currency type | No |
| CARD_NUMBER | varchar(20) | Customer's card number | Yes |
| VALIDITY | date | Customer's Card Validity | No |
| CVV | int | Customer's Card Verification value | Yes |
| ACTIVE_FLAG | char(1) | Flg to check if customer is active | No |
| AUTO_RENEW_FLAG | char(1) | Flag for auto renewal of subscription | No |
| PAYMENTMETHOD | varchar(20) | Method of payment | No |
| DISCOUNT | int | Discount on the basis of promo code | No |
| TOTALPRICE | numeric(18, 2) | Total payment amount after discount(if applies) | No |
| SUBSCRIPTION_END_DATE | datetime | Subscription end date | No |
| BILLABLE_MONTH_TILL_DATE | int | Subscription Duration | No |

Fig. 11. Data Dictionary of Transaction Table

## VII. GENERAL DATA PROTECTION REGULATION (GDPR) AND ACCESS CONTROL

The General Data Protection Regulation is a European Union-wide framework that came to action in 2018 [3]. Although Netflix is an American company, to expand its business in the European it has adhered to the rules of GDPR. As per GDPR, all the personal data like Name, contact information, address of a subscriber, card details that would directly or indirectly reveal the information of a person should not be revealed. So, obeying the rules, all the columnar information that reveals the personal information are masked at the data base level.

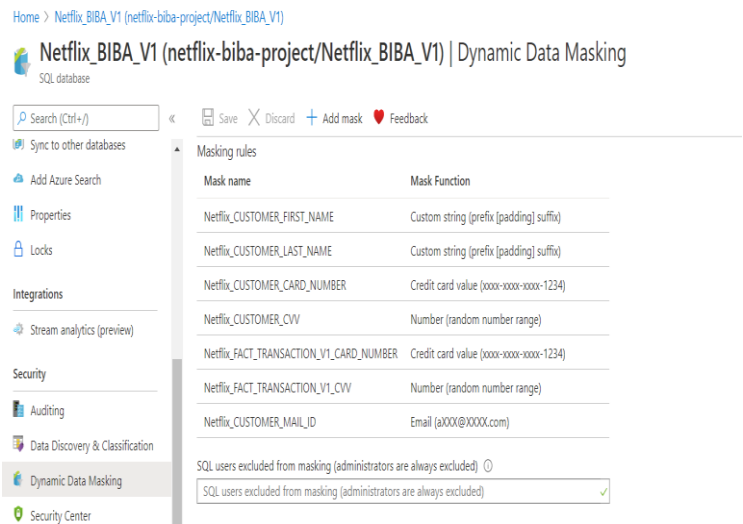The below picture shows masking of columns applied on Azure:



Fig. 12. Data Masking

Also, to keep the customer's trust intact the access for the data servers is made strictly accessible only to the limited employees, say only to the project team in the analysis.
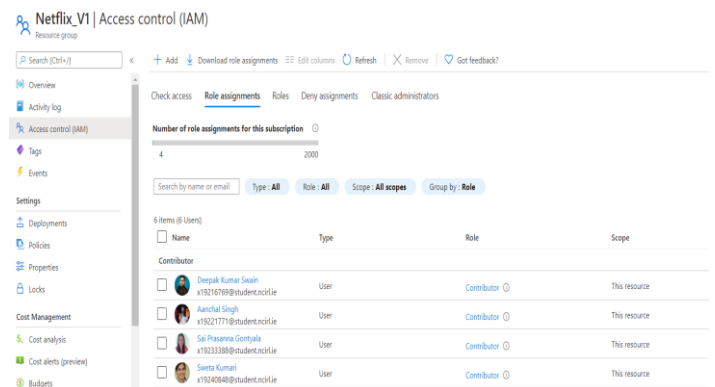
The below picture shows the access control.



Fig. 13. Access control of Resource Group

## VIII. BUSINESS RULES

Implementation of business rules helps in having refined data which in turn helps in creating high quality reports. In the analysis, we have maintained the data considering the below business rules:

- Every customer in the system should be tagged to a subscription plan else the customer is marked inactive in the system.

- Only the shows or movies that have an active contract with Netflix is only maintained in the system else the status of it is changed to inactive.

- Every promocode that is released by Netflix that has helped in converting a lead to customer through any means of marketing should be recorded in the system against its active period.

- Every successful transaction should be recorded in the system.

- If the subscriber chooses Auto Renewal of the subscription it must to maintain the card details of the subscriber.

## IX. DATA INTEGRITY

It is important to have data integrity in the system for an effective analysis as all the entities at any point of time would it holding only the true information and helps all the entities in the system to stay intact.

In order to achieve this, foreign key relationship is maintained on the interconnected columns so that the data is true and intact. Null Constraints are defined on Few fields like Customer ID, Billing Date, Payment method that would help having error free data that is aligned with the business process.

## REFERENCES

[1] https://www.kaggle.com/shivamb/netflix-shows
[2] A Comparative Analysis of Entity-Relationship Diagrams by Il-Yeol Song, Mary Evans, E.K. Park
[3] https://gdpr-info.eu/