Marketplace Feature Table

Your**Client ComZ** is an ecommerce company. The company wants to focus on targeting the right **customers**  with the right products to increase overall revenue and conversion rate.

To target the right customers with the right products, they need to build an ML model for marketing based on user interaction with products in the past like number of views,  most viewed product, number of activities of user, vintage of user and others.

**ComZ has** contacted the Data Science and Engineering team to use this information to fuel the personalized advertisements, email marketing campaigns, or special offers on the landing and category pages of the company's website.

You, being a part of the data engineering team, are expected to “**Develop input features”** for the efficient marketing model given the **Visitor log data** and **User Data**.

1. **Visitor Log Data**– It is a browsing log data of all the visitors and the users. This table contains the following information:

|  |  |
| --- | --- |
| WebClientID | Unique ID of browser for every system. (If a visitor is using multiple browsers on a system like Chrome, Safari, then there would be a different web clientid for each browser). The ID remains consistent unless the user clears their cookie. |
| VisitDateTime | Date and time of visit. There are two different formats for DateTime.   * One is in datetime format “2018-05-07 04:28:45.970” * Another one is in unix datetime format “1527051855673000000” |
| ProductID | Unique ID of product browsed/ clicked by the visitor |
| UserID | Unique ID of the registered user. As expected, this is available for registered users only, not for all visitors. |
| Activity | Type of activity can be browsing (pageload) or clicking a product |
| Browser | Browser used by the visitor |
| OS | Operating System of the system used by the visitor |
| City | City of the visitor |
| Country | Country of the visitor |

**2. User Data** – It has registered user information like signup date and segment.

|  |  |
| --- | --- |
| UserID | Unique ID of the registered user. |
| Signup Date | Date of registration for the user |
| User Segment | User Segment (A/B/C) created based on historical engagement |

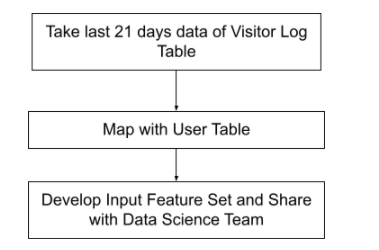
Now based on the above two tables, you need to create an **input** **feature set** for the Marketing Model.

**3. Input Feature table:**

|  |  |
| --- | --- |
| UserID | Unique ID of the registered user |
| No\_of\_days\_Visited\_7\_Days | How many days a user was active on platform in the last 7 days. |
| No\_Of\_Products\_Viewed\_15\_Days | Number of Products viewed by the user in the last 15 days |
| User\_Vintage | Vintage (In Days) of the user as of today |
| Most\_Viewed\_product\_15\_Days | Most frequently viewed (page loads) product by the user in the last 15 days. If there are multiple products that have a similar number of page loads then , consider the recent one. If a user has not viewed any product in the last 15 days then put it as **Product101.** |
| Most\_Active\_OS | Most Frequently used OS by user. |
| Recently\_Viewed\_Product | Most recently viewed (page loads) product by the user.  If a user has not viewed any product then put it as **Product101.** |
| Pageloads\_last\_7\_days | Count of Page loads in the last 7 days by the user |
| Clicks\_last\_7\_days | Count of Clicks in the last 7 days  by the user |

**Process to create Input Feature:**

When ComZ does a targeting campaign, It follows the below process.



In the current case, you are supposed to generate an input feature set as on 28-May-2018. So, the visitor table is from 07-May-2018 to 27-May-2018.

*As a Data Engineer Creating ETL Pipeline would definitely be appreciated and provide you the added advantage in interviews, Your effort should be to build ETL* *Pipeline such that passing the information of user data and log data, It can generate the input feature table automatically*