## **Marketplace Feature Table**

Project Document

4th July 2021

Contents

[Business Case 3](#_Toc76319593)

[Project Description 3](#_Toc76319594)

[Process Flow 3](#_Toc76319595)

[Environment Setup 4](#_Toc76319596)

[Approach 4](#_Toc76319597)

[Data Pre Processing 5](#_Toc76319599)

[Tools Used 6](#_Toc76319600)

[Main Flow 6](#_Toc76319601)

[Project Code 7](#_Toc76319602)

# 

# 

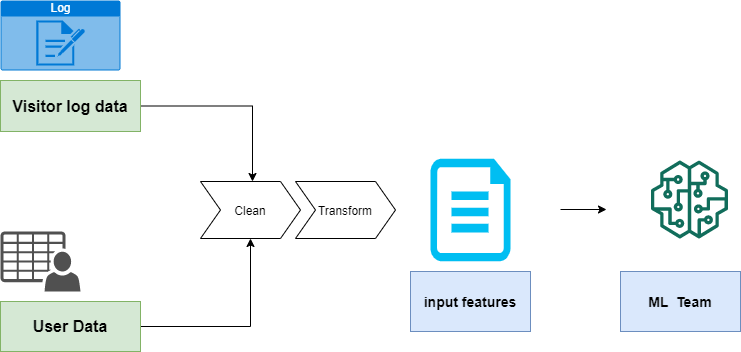
## Business Case

**Client ComZ** 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.

## Project Description

The aim of the project is to “**Develop input features”**  for the efficient marketing model given the **Visitor log data** and **User Data** to fuel the personalized advertisements, email marketing campaigns, or special offers on the landing and category pages of the company's website.

## Process Flow

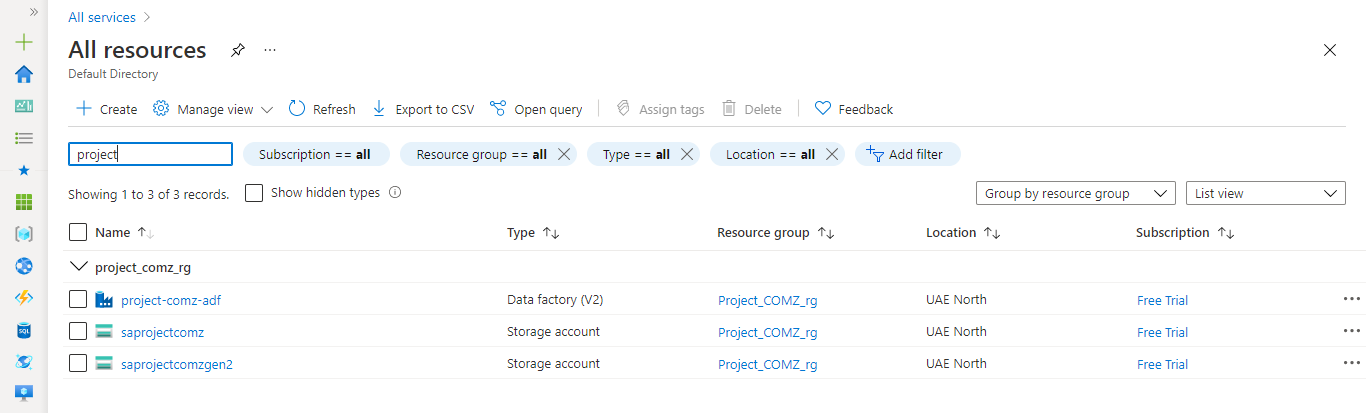


## 

## 

## Environment Setup

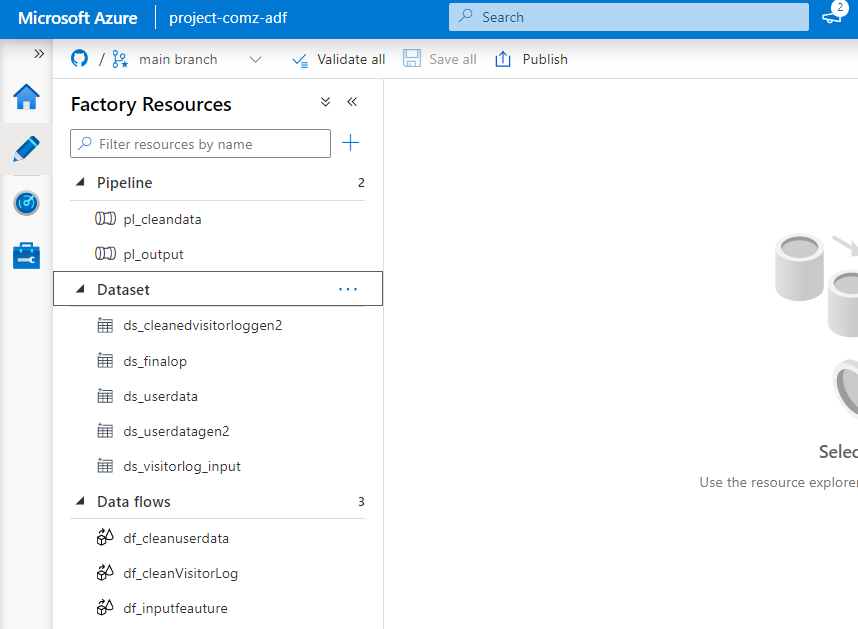
* Creating Azure Data Factory
* Creating Azure Storage Account
* Create Azure Data Lake Gen2



## Approach

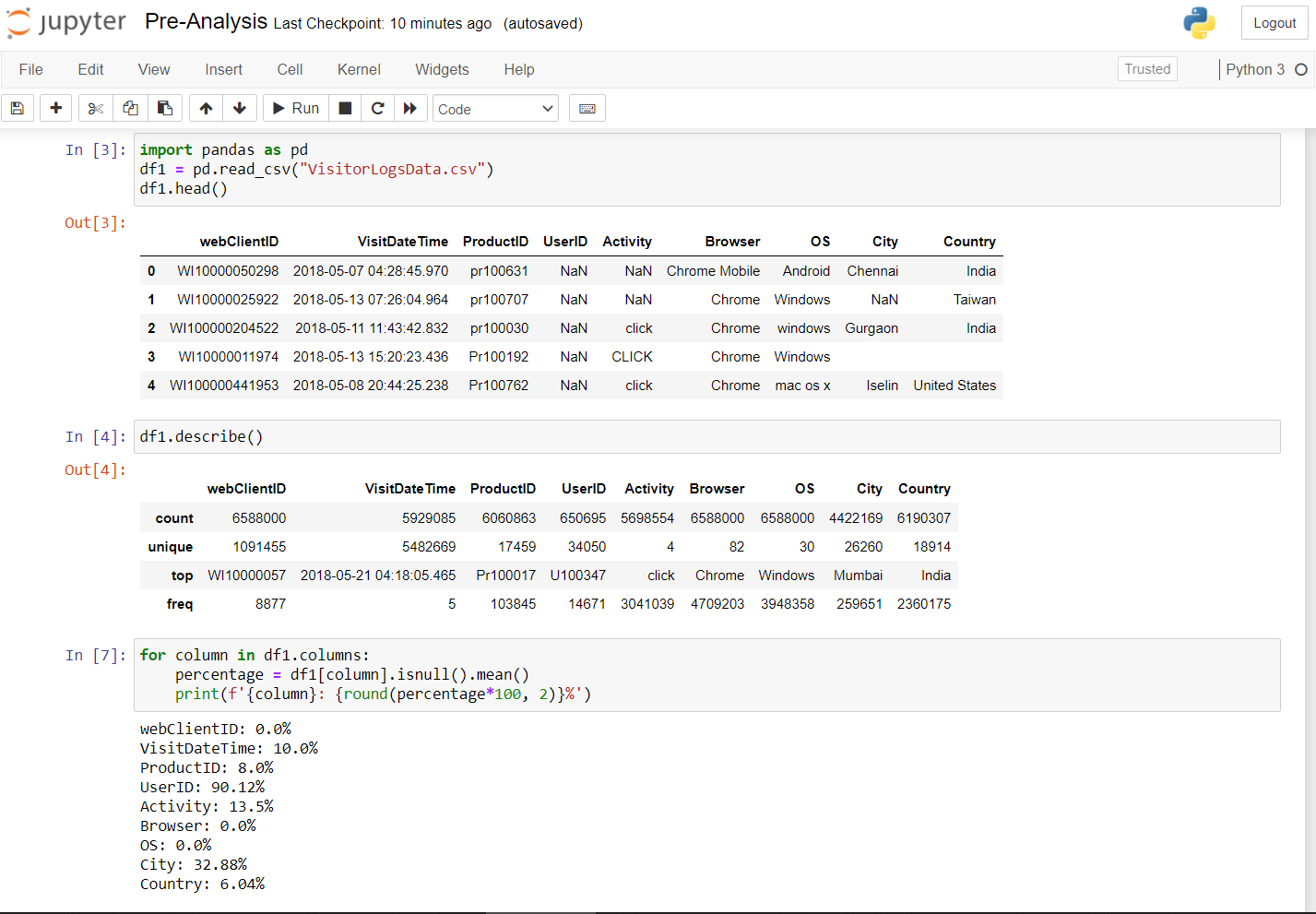
## Approach was to implement the given ETL pipeline using Azure Data Factory .The whole cleaning and processing were carried out using Mapping Data Flow in Azure Data Factory . Steps Involved were

* Pre-Analysis of the input files were carried out in Python
* Plan was to implement the entire ETL pipeline in Azure Cloud
* Created a free subscription in Azure
* Created the resource group for the project
* Created two storage accounts - Blob storage for storing the raw files and Gen Lake 2 storage for storing the processed file
* Created Azure Data Factory
* Created Linked services to connect to data sources ,one for Blob storage and other of Gen2 storage
* Three Data flows were made 2 for data cleaning and one main flow
* Two pipelines were created one for invoking the cleaning dataflows other for invoking the main data flow activity

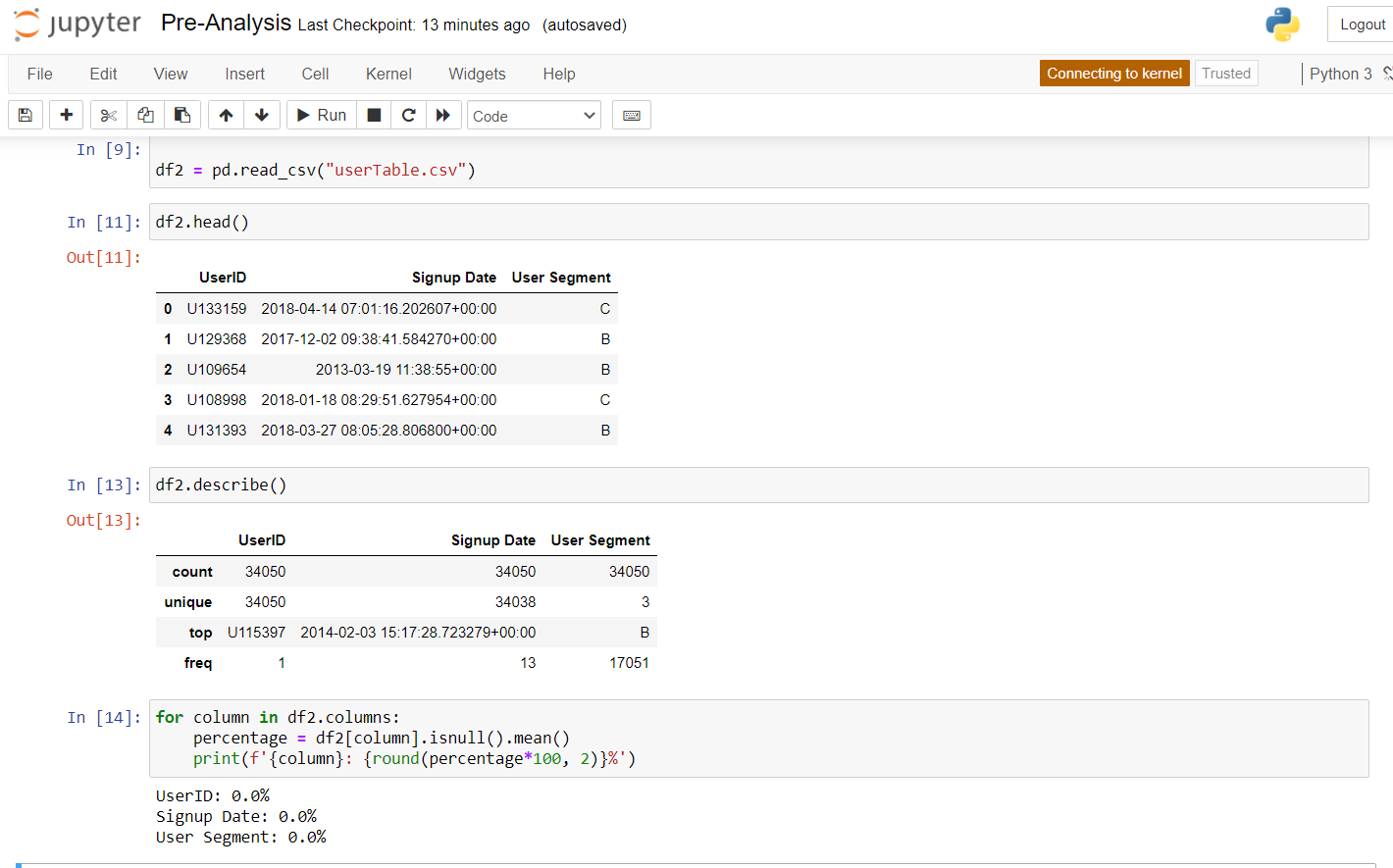


## Data Pre-Processing

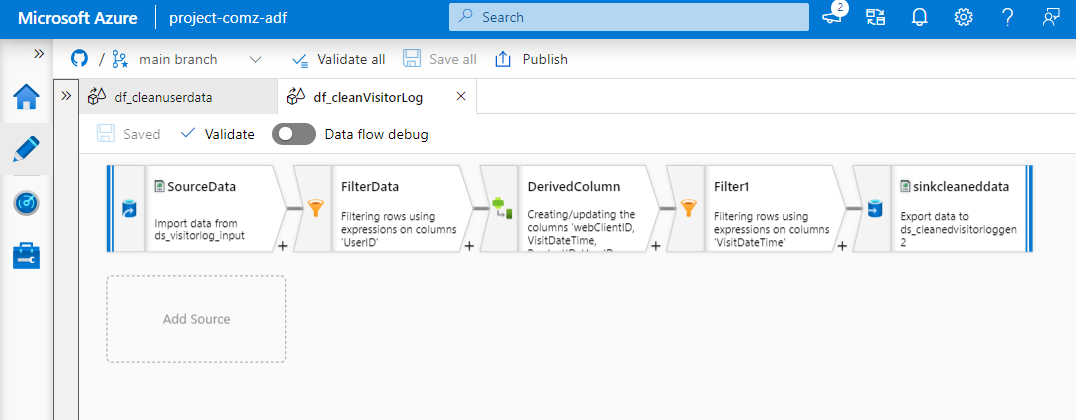
Initially analysis was carried out using python pandas , identifying the non nulls



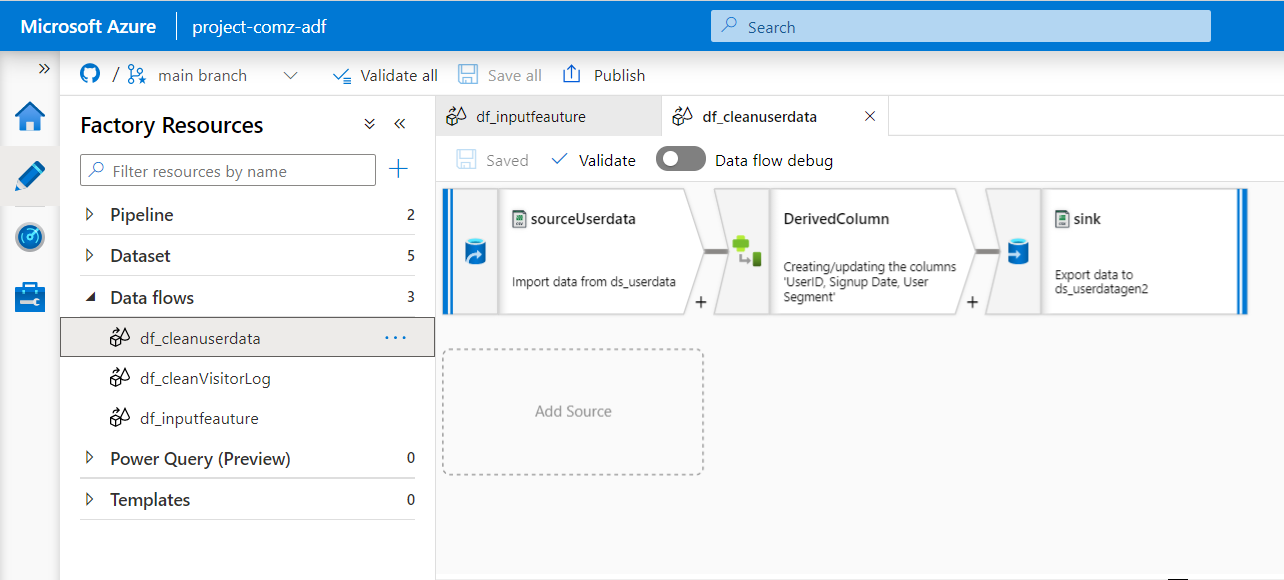
Visitorlog data Analysis



User Data Analysis



Using Data Flow the visitor log data was cleansed and written to using uppercase ,trim function also by filtering the non-null User ID ,and fetching the required dates

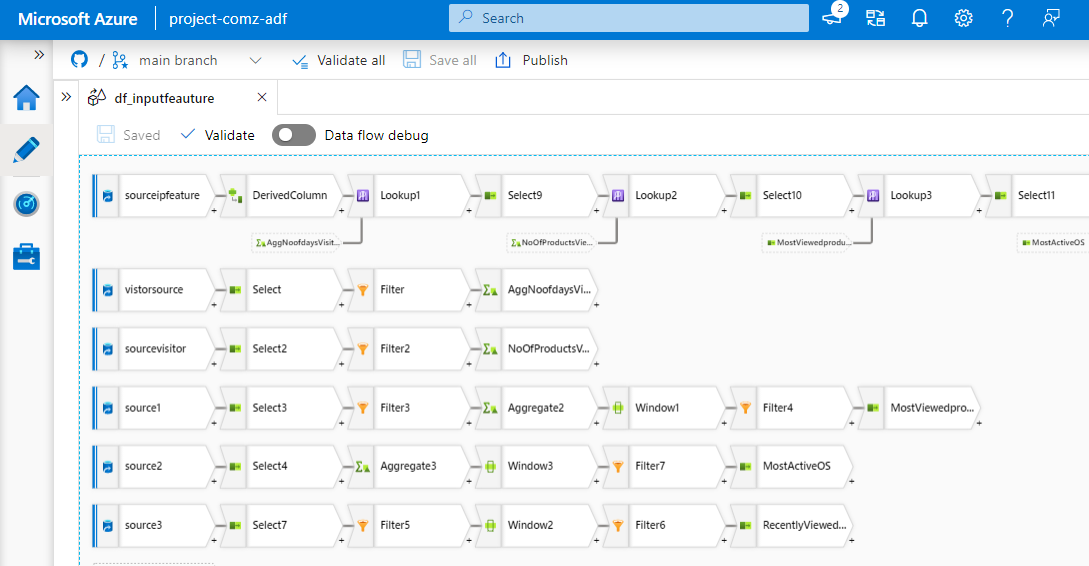


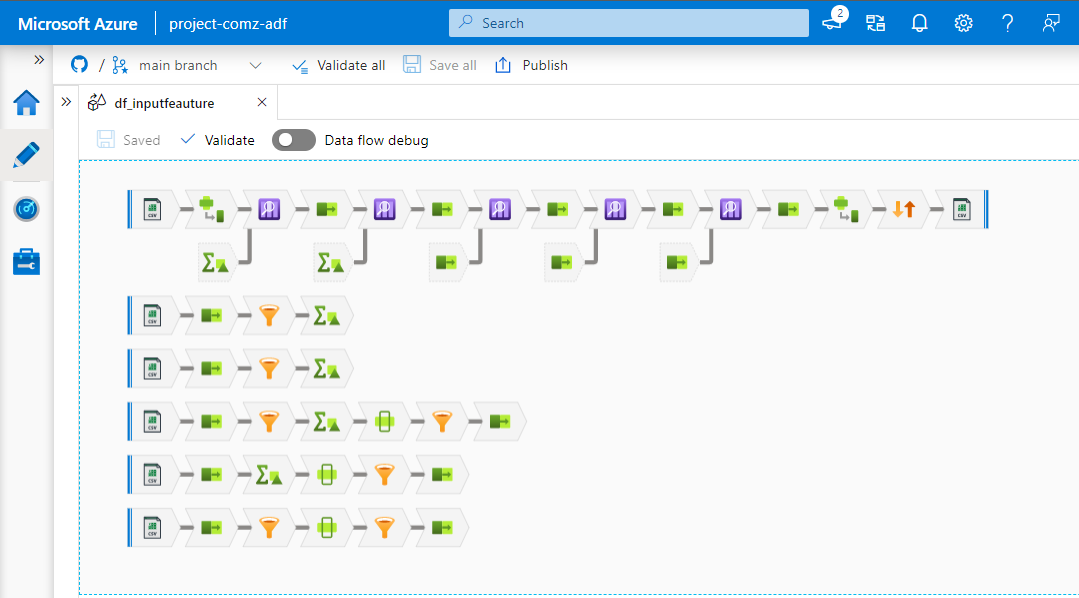
Using data flow the user data was cleaned using uppercase and trim function

## Tools Used

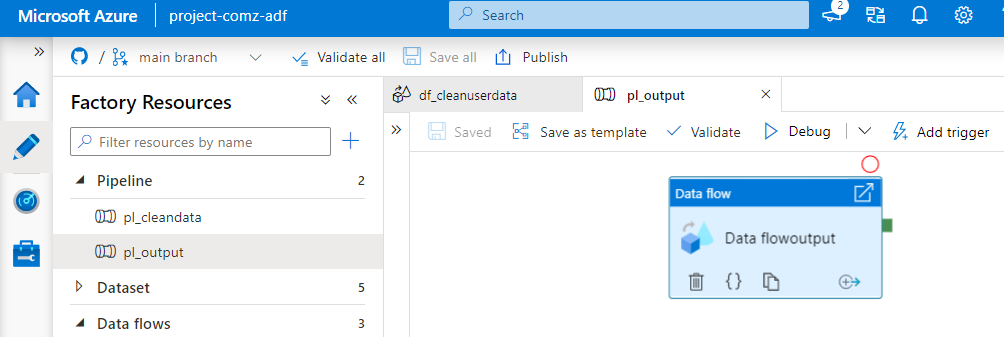
* Python(pandas) - Initial analysis
* Azure Data Factory
* Azure Mapping DataFlow
* Azure Storage account

## Main Flow -Solution





df\_inputfeature - Mapping Data Flow



ADF Data Flow activity for input feature

## Project Code

<https://github.com/sanjusuresh254/project_comz.git>