Group 40:

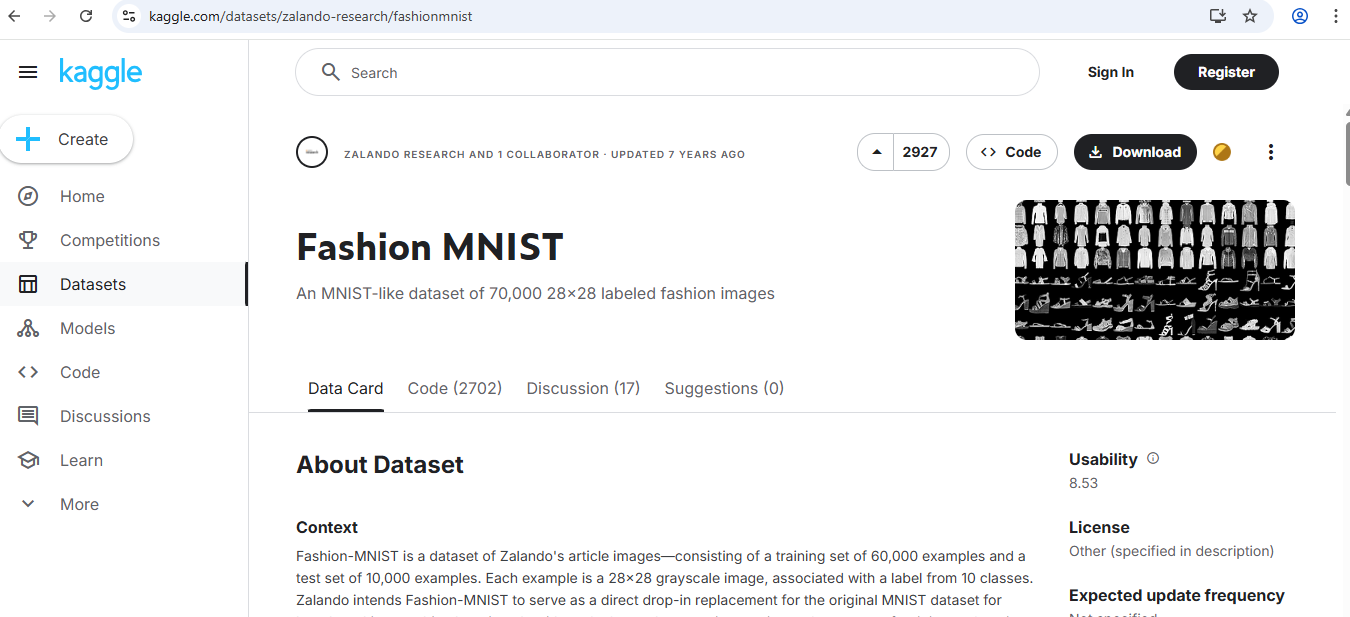
|  |  |
| --- | --- |
| 2023aa05995 | Akshara.B |
| 2023ab05053 | Chandhira sekaran N |
| 2023aa05501 | Deepak C C |
| 2023aa05996 | Rajesh J |

Description of the work done.

M1 – EDA: Exploratory Data Analysis

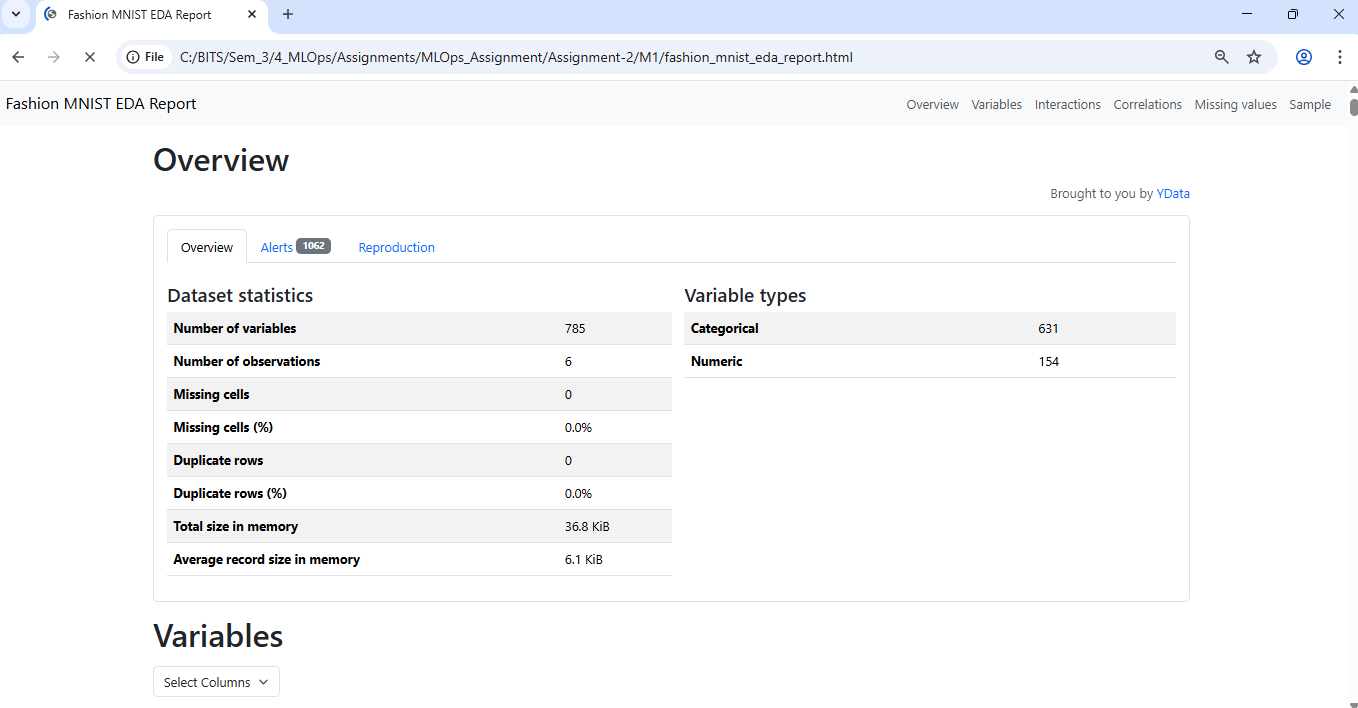
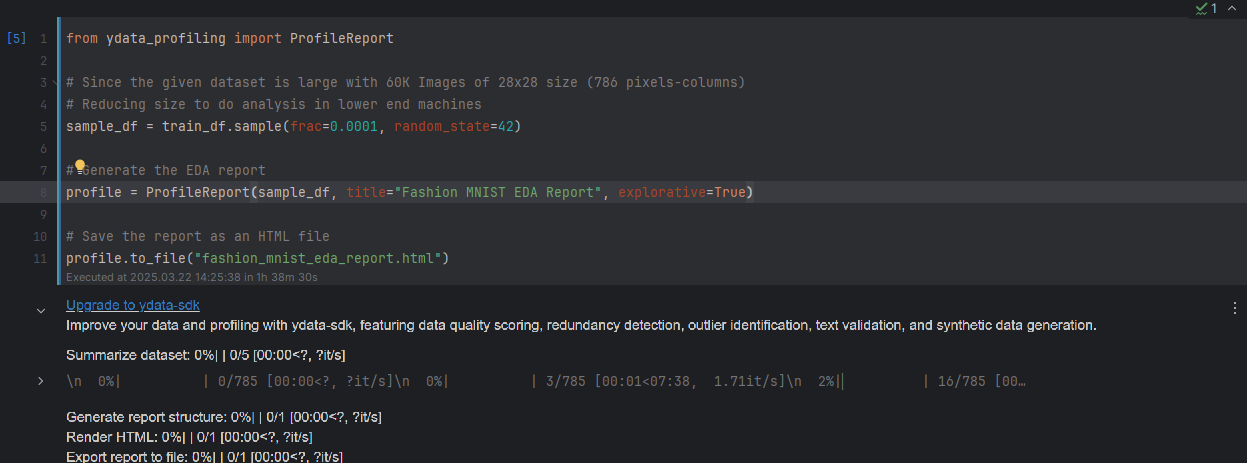
**Fashion MNIST Dataset:**

The **Fashion MNIST Dataset** was used for the analysis from Kaggle - <https://www.kaggle.com/datasets/zalando-research/fashionmnist> which is an Image dataset of 70,000 grayscale images with a pixel size of 784 (28 X 28 images) columns.

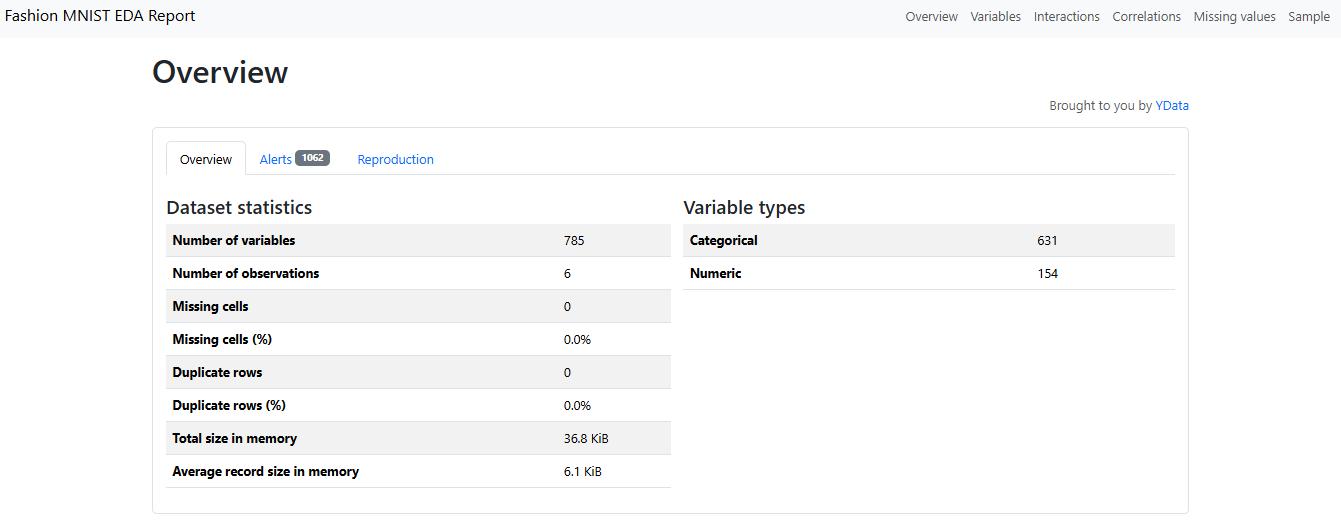
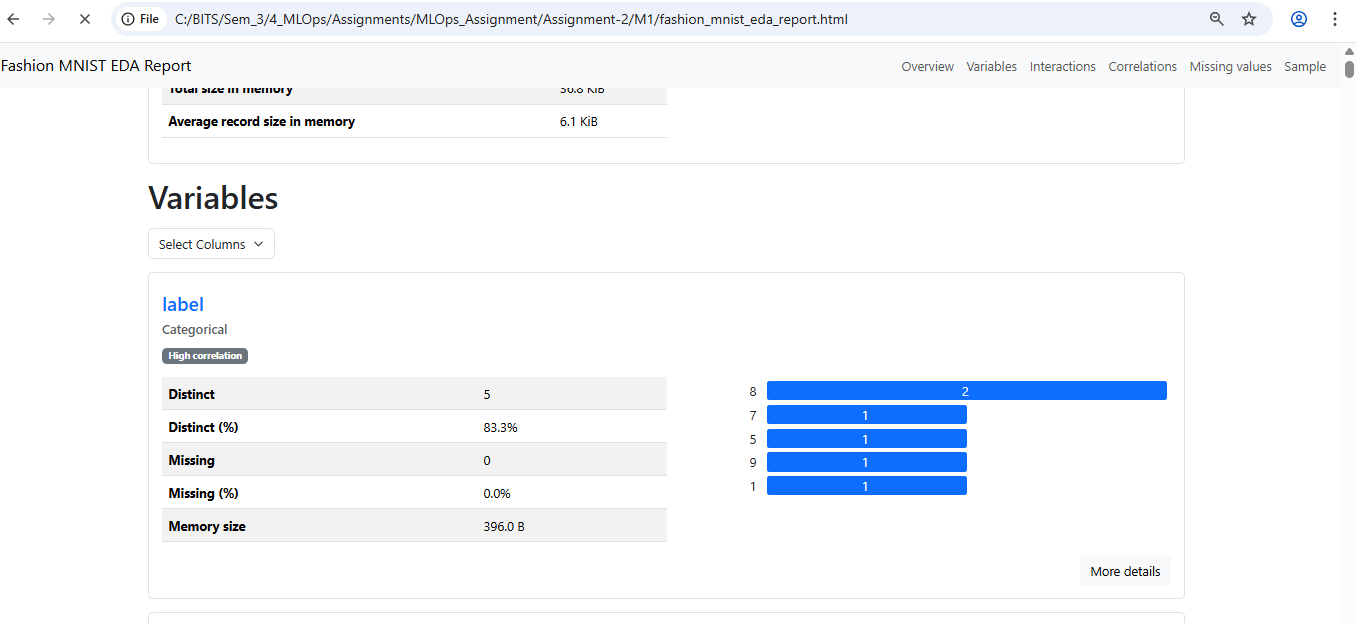
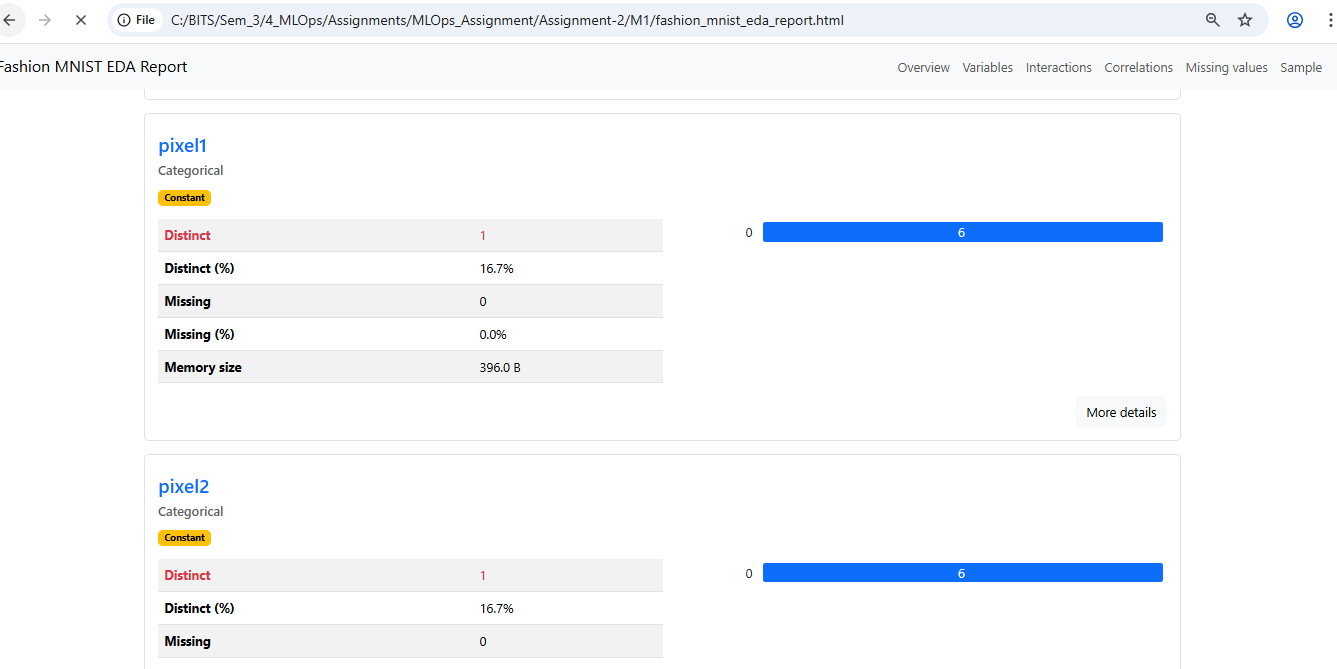
Labels values ranging from 0 to 9 was present as the first column in the dataset which is used to classify the images as shown:  
0 T-shirt/top, 1 Trouser, 2 Pullover, 3 Dress, 4 Coat, 5 Sandal, 6 Shirt, 7 Sneaker, 8 Bag, 9 Ankle boot  


As this is huge dataset with 70K images, only a sample of 0.01% was used for the data analysis part.   


**Profiling Using Ydata Profiler:**

As part of this task, we have automated Data Analysis using Pandas Profiling (new name Ydata-Profiler) and have generated the report named “***fashion\_mnist\_eda\_report.html***” as shown:  
  
  
Since the size of the generated report exceeds 300MB in size, it will be uploaded in GDrive with the link shared here – <https://drive.google.com/file/d/11CQ2wgKswHPrDm6soMpJ8RZE7GFWDS-K/view?usp=sharing>

**M1 Conclusions:**

As shown in the overview of the EDA report with 784 features and the first column as the Label column:  
  
The report also shows the different analysis like Missing Values, Distinct Values, Correlation etc among many observations.  
For example, the output column named “**Label**”, shows that’s it’s a Categorical Field, with distinct 5 values. No missing values and the memory size it constitutes. Additionally, it also shows a Plot of distribution bar-chart for visual correlations.  
  
For another feature column named “Pixel1” here are the screenshots which implies a constant or in other works no-correlation to other fields. Additionally, it also shows a Plot of distribution bar-chart for visual correlations.  
  
For columns like Pixel14 and Pixel15, it shows higher distinct values and thereby higher correlation. Additionally, it also shows a Plot of distribution bar-chart for visual correlations.  
