

MACHINE LEARNING

This part of the project uses the data we have obtained on the preceding parts and processes the data to earn insights from it and create Machine Learning models. The insights will be used to produce a user behavior and to determine the strategical approach of our personalized advertisement product.

- **visualizations_output_tasks**

The folder contains the deliverables/outputs of each task.

- **machine_t1.ipynb:**

Objective of the task: Detect all overall number of TV devices and categorize them based on how long they watch specific TV channels in specific time period.

Using the CSV file (with the entire data store in the Snowflake database) containing each user session per channel, the program evaluates the time spent of each user on each channel from 8.00 to 14.00.

Apart from accomplishing the main goal, the set of algorithms can determine:

- a) Average time users spent per channel.
- b) Frequency count per channel/get most popular channel among users.
- c) Total unique users per channel.
- d) Get user's location (country, city, postal code) with its IP address.

- **machine_t2.ipynb:**

Objective of the task: Work on an algorithm that will help predict amount of traffic (TV viewership) on the time period from 2PM till 11:59PM with single source.

Using the CSV file containing each user session, the program will filter daily traffic from 14.00 up to 23.59 to compare them with the same day and time on different weeks, to prepare a prediction using Support Vector Machine on how the traffic would be like on the same days of the next weeks.

- **machine_t3.ipynb:**

Objective of the task: Use an anomaly detection algorithm to predict and spot unusual behavior by using the “Duration per TV session” feature with each channel. We identify unexpected events in data set. With the algorithm we can:

- Classify each observation as an outlier or non-outlier.

- Display a visualization that highlights the regions where the outliers fall.