

How Tavant Decoded Television Viewership Data To Enhance The Offerings Of A Global Content Service Provider



Ravi Peravali

Vice President - Media Practice,
Tavant

Harsha Kommanapalli

Director - Machine Learning,
Tavant

Sunil Kumar Yadav

Lead Data Scientist,
Tavant



Executive Summary

The customer, a global content service provider, is a pioneer in home entertainment, with millions of customers tuning in daily to discover unique experiences on television. They also have an array of solutions for television service providers, content and streaming services, advertisers, etc. Considering the vast number of consumers engaging with their platform, the client had access to a wealth of consumer engagement and viewership data. They wanted to partner with a technology powerhouse with strong media domain expertise to focus on viewership data and employ intelligence to conceptualize innovative solutions for advertisers and other clientele.

Business Overview

The client offered comprehensive data products across the multi-channel ecosystem. Their dataset integrated demographic attributes, ad airings, and digital devices data into TV viewership data. The data was further enriched by leveraging partners such as Experian, LiveRamp, and others.

In order to get ahead in their journey of using analytics to improve business, the client equipped Tavant with anonymized consumer behavior and content engagement data for 3 million US households along with program data. Each of these programs was mapped to their right set of genres.

This article provides a detailed description of the data sets provided by the client, the technology employed to process this data and use-cases that were developed by Tavant to benefit the business and customer pool of the client.

Data Trail

The following data points were made available to Tavant via an Amazon Web Services S3 Bucket:

- ▶ 3.4 billion rows of return path data derived from a 30-day watch period from 3 million households
- ▶ 28 million rows of genre data after mapping every program to a fixed set of genres

Viewership Data

Viewership data is a collection of watch-events. Each row of viewership data corresponded to a program watch-time of a household. Some of the parameters captured were:



Household ID corresponding to the watch-event



Actual program watch duration by the household



Program watched



Location of the household



Network, channel airing the program



View event type - linear (live) or recorded



Scheduled time and duration of the program

Genre Data

The genre data mapped every program to a sequence of associated genres. Programs and genres were associated with unique IDs. A program could have multiple genre associations, and this was captured across multiple rows of genre data.

Data Pre-Processing: OLAP Cube

The total size of the compressed data was around 250 GB. Analyzing this data each time the user interacts with the dashboard was impractical, slow, unresponsive, and would not have met the purpose of having an interactive dashboard. Tavant created an OLAP cube by rolling up on several dimensions until the program level. The OLAP cube had over 29 million entries (compared to 3.4 billion entries in the viewership data).

In this context, OLAP cube, created using SQL, refers to a table that holds several summarized metrics corresponding to multi-dimensional groups. The term 'OLAP cube' is used to refer to pre-summarized tables and not any commercial OLAP solution.

Exploratory Data Analysis

Once the data was imported, we performed exploratory data analysis on the same to find any interesting patterns. Some of the findings reported were:

- ▶ Watch Hours – weekday vs. weekend
- ▶ Watch – Day-Part Wise
- ▶ Network – View Share
- ▶ Top Watched Networks
- ▶ Number of unique households at different watch hours
- ▶ Density Map: Average Watch Percent vs. Watch Hour - A Kernel Density map showed the relationship between program watch percentage and the hour at which it was watched. Program watch-percent is the percentage of a program that is watched by the user.

Applications Designed

Tavant helped the client map the journey of every household by analyzing their various touchpoints. Three use-cases that would be relevant to TV networks and marketing companies using the given datasets, were identified.



Enable broadcasters and TV networks to get viewership metrics of their programming content

This was demonstrated by an interactive dashboard that would capture viewership metrics across geographies, along different times of the day, across channels and genres.



Identify and analyze similarities in viewer behavior at a household level

This will involve clustering by viewing patterns of households



Analyze patterns across geographies and genre affinities

This will involve clustering based on hours spent on different genres



Interactive Dashboard Application

The interactive dashboard application, a web-based GUI application allowed users to explore the affinities of different locations in the US towards various TV networks, genres, dayparts, and weekdays. The application painted a gradient-map of watch-hours across different geographies (DMAs on the US map) allowing the user to understand relative viewing habits.

The dashboard was primarily divided into two sections:

- ▶ User Controls, on the left panel, where users could choose what they want to see
 - Top 10 networks
 - Top 10 genres
 - Days of the week
 - Day time slider
 - Normalized calculation
- ▶ A choropleth map layered on US Geography on the right panel - a gradient map corresponding to the left-panel based on watch-hours

Grouping Similar Households

KMeans clustering was used to group audiences/households into 'K' groups. Each group would have similar watch behaviors such as:

- Times of the day they watch TV
- TV genres they like to watch
- Frequency of 'Live Watch' v/s 'Recorded and Watched Later'
- TV networks they usually watch

Tavant built a total of 31 features from the dataset for each household and the clustering was done in 3-dimensional space. Clustering analysis could help understand groups of users who were very similar in their watch behavior. This analysis was extended to create effective marketing strategies based on watch-pattern and household similarities.

Geography-Genre Affinity

In this use case, Tavant identified geographies that have similar affinities towards different genres. The locations were plotted such that similar geographies were marked with same colors on the US map.

To identify geographies that are viewing similar content, Tavant arrived at different viewing patterns (a set of watch-hours for different genres) among the total population.

The way ahead

Data is key to the sustenance and growth of television. With the intelligence derived from the volume of information media companies can obtain in today's age, viewers will get to see better content, accurate and personalized recommendations amongst others. Television programming and scheduling will also be carried out in a planned fashion, backed with statistics. Advertisers will tailor their offerings as they have increased access to information on their viewers and attribution related to their spend.

Tavant has a well-established practice focused on working with Media and Ad Tech companies, to develop innovative bespoke solutions. Tavant's media expertise has grown to accommodate the range of services required across the media sub-verticals. Having the breadth to cater to sub-vertical needs such as Broadcasting, Publishing, New Media, Sports, and Ad Tech, Tavant helps Media and Entertainment companies in driving digital transformation by providing strategic solutions in line with new business models, emerging technologies, and changing audience demands.