

STC Jawwy

Jawwy dataset

The dataset consists of meta details about the movies and tv shows as genre. Also details about Users activities, spent duration and if watching in High definition or standard definition. You have to analyse this dataset to find top insights, findings and to solve the four tasks assigned to you.

...	↑↓	...	↑↓	...	↑↓	user_i...	...	↑↓	program_name	...	↑↓	duration_se...	...	↑↓	progra...	...	↑↓	...	↑↓	...	↑↓	program_
	0		1		42882			26138	100 treets					40	MOVIE			0		0		Drama M
	1		3		42876			7946	Moana					17	MOVIE			0		0		Animation
	2		4		42957			7418	The Mermaid Princess					8	MOVIE			0		0		Animation
	3		5		42942			19307	The Mermaid Princess					76	MOVIE			0		0		Animation
	4		7		42923			15860	Churchill					87	MOVIE			0		0		Biography

Rows: 5

Expand

...	↑↓	date_	...	↑↓	user_i...	...	↑↓	program_name	...	↑↓	duration_se...	...	↑↓	progra...	...	↑↓	...	↑↓	...
	0	1970-01-01T00:00:00.000			26138			100 treets					40	MOVIE			0		
	1	1970-01-01T00:00:00.000			7946			Moana					17	MOVIE			0		
	2	1970-01-01T00:00:00.000			7418			The Mermaid Princess					8	MOVIE			0		
	3	1970-01-01T00:00:00.000			19307			The Mermaid Princess					76	MOVIE			0		
	4	1970-01-01T00:00:00.000			15860			Churchill					87	MOVIE			0		

Rows: 5

Expand

i.	...	↑↓	date_	...	↑↓	duration_seconds	...	↑↓	season	...	↑↓	episode	...	↑↓	series_title	...	↑↓
count			1048575			1048575			1048575			1048575			1048575		
mean			1970-01-01T00:00:00.000			1230.9572124073			1.3421386167			6.1579524593			0.0120592232		
min			1970-01-01T00:00:00.000			2			0			0					
25%			1970-01-01T00:00:00.000			52			0			0					
50%			1970-01-01T00:00:00.000			119			1			1					
75%			1970-01-01T00:00:00.000			1328			1			9					
max			1970-01-01T00:00:00.000			1461329			23			282					
std			null			6821.0578889847			2.1040953822			12.2201546952			0.1091503996		

Rows: 8

Expand

index	...	↑↓	0
date_			False
user_id_mapped			False
program_name			False
duration_seconds			False
program_class			False
season			False
episode			False
program_desc			False
program_genre			False
series_title			False
hd			False
original_name			False

Rows: 12

Expand

Task 1

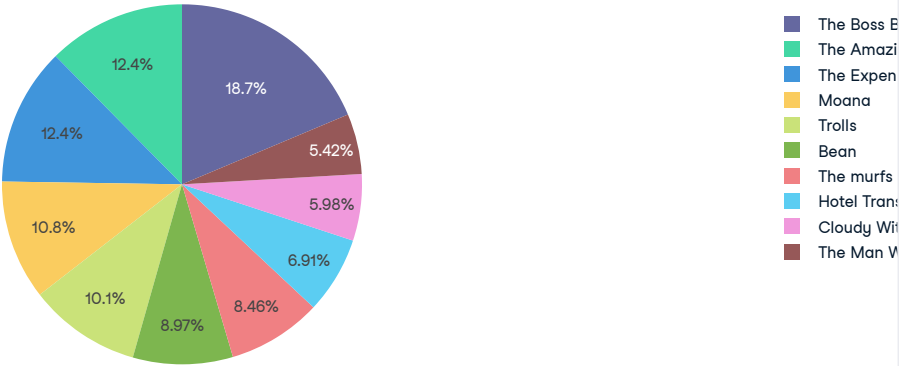
You are required to work on task one to study and HD flag for available dataset

...	↑↓	program_name	...	↑↓	program_cl...	...	↑↓	No of Users who Watched	...	↑↓	No of w...	...	↑↓	Total watch time in
	0	The Boss Baby			MOVIE			3389			24047			
	1	The Amazing pider-Man			MOVIE			1011			2877			
	2	The Expendables			MOVIE			853			2119			
	3	Moana			MOVIE			2173			8081			
	4	Trolls			MOVIE			2613			13793			
	5	Bean			MOVIE			949			3617			
	6	The murfs			MOVIE			867			3132			
	7	Hotel Transylvania			MOVIE			491			1947			
	8	Cloudy With a Chance of Meatballs			MOVIE			683			2076			
	9	The Man With The Iron Fists			MOVIE			707			2505			
	10	Salt			MOVIE			563			1082			
	11	Unbroken			MOVIE			625			1429			
	12	ParaNorman			MOVIE			614			1746			
	13	Youm Maloosh Lazma			MOVIE			1131			2278			
	14	Ferdinand			MOVIE			1278			6817			
	15	White Chicks			MOVIE			307			916			

Rows: 35

Expand

top 10 programs in total watch time in heures

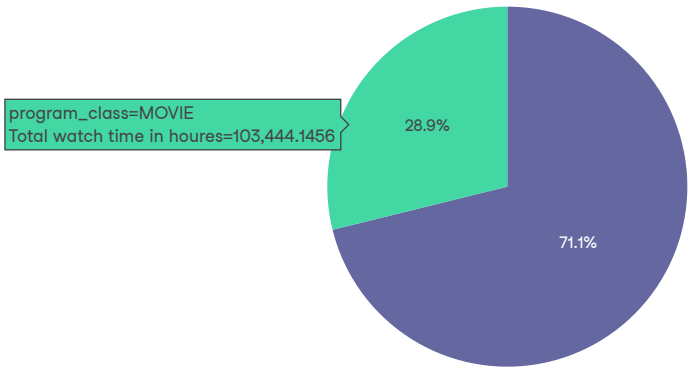


in...	...	↑↓	program_class	...	↑↓	No of Users who Watched	...	↑↓	No of watches	...	↑↓	Total watch time in hours
	0		SERIES/EPISODES			3901			560174			
	1		MOVIE			11355			488401			

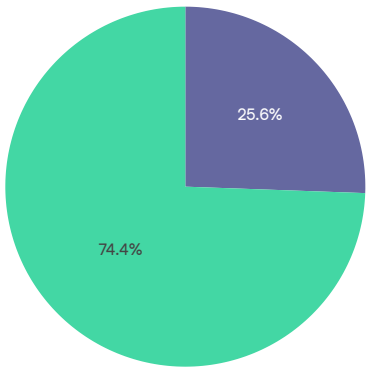
Rows: 2

Expand

Total duration spent by program_class



Total Users watching by program_class

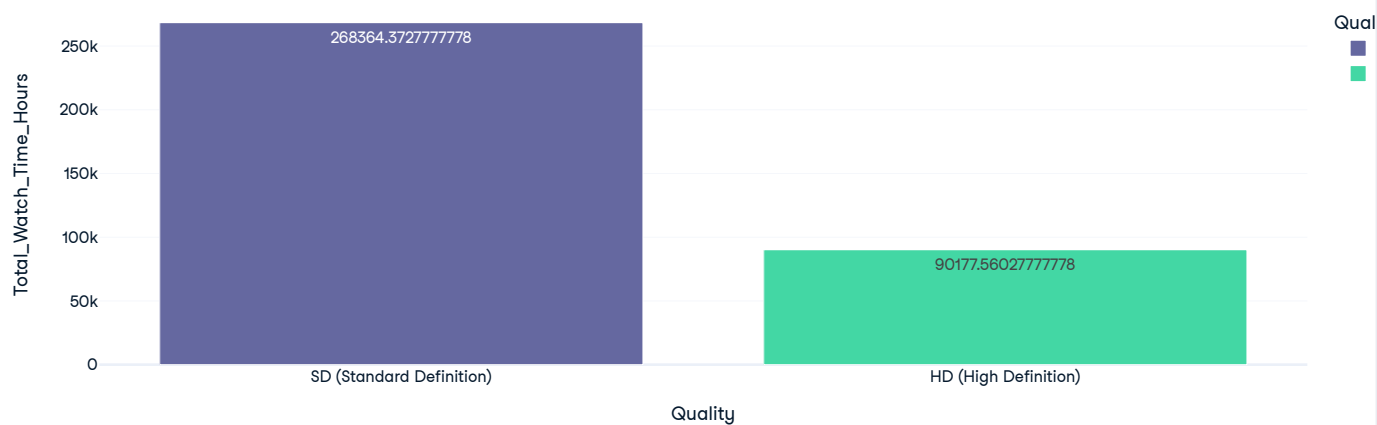


in...	...	↑↓	H...	...	↑↓	Unique_Users	...	↑↓	Total_Views	...	↑↓	Total_Watch_Time_Hours	...	↑↓	Quality
	0			0		6728			643539			268364.3727777778			SD (Standard Definition)
	1			1		11000			405036			90177.5602777778			HD (High Definition)

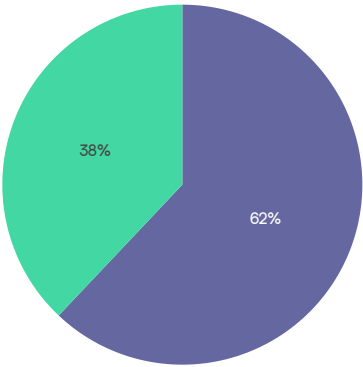
Rows: 2

Expand

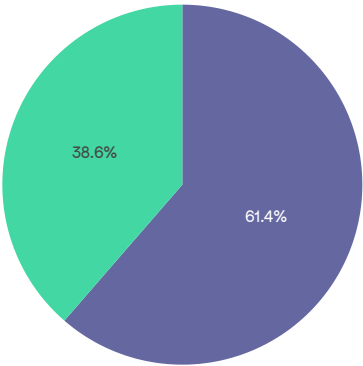
Total Watch Time (Hours) by Video Quality



Unique Users by Video Quality



Total Views by Video Quality



Although a larger number of users engage with HD content, SD dominates in both total views and total watch time. This could indicate:

Limited access to high-speed internet or HD-capable devices for many users.

Background streaming behavior (e.g., kids watching animation or series on loop) where video quality is less important.

Understanding this behavior is crucial for content delivery optimization and platform infrastructure planning.