

APPLE PRODUCTS AND APP USAGES



Contents

1. Business Challenge / Requirement	03
2. Goal of the project	03
3. Project Architecture.....	04
4. Dataset Explanation and Schema	05
5. Problem Statement / Tasks.....	10
6. Code Templates.....	11
6.1 Data Processing.....	12
6.1.1 Conversion of Raw data into Processed Data.....	12
6.1.2 Process Data.....	13
6.2 Hive and Sqoop.....	16
6.3 Hadoop Code.....	23
7. Output Screen.....	24
8. Business Benefits/ Conclusion.....	26
9. Further Enhancements/Recommendations	26
10. References/Bibliography.....	26

1. Business Challenge / Requirement

A Tech product company called Apple is facing challenges in understanding the usage of its internal audio, video, and news applications in comparison to external applications by users of Apple devices such as iPhone, iPad, Macbook, etc. So the company wants to take the help of Data Analytics using Big Data Ecosystem to analyse the data of users in order to increase its internal app usage and prevent users based on their age group from opting to external party apps.

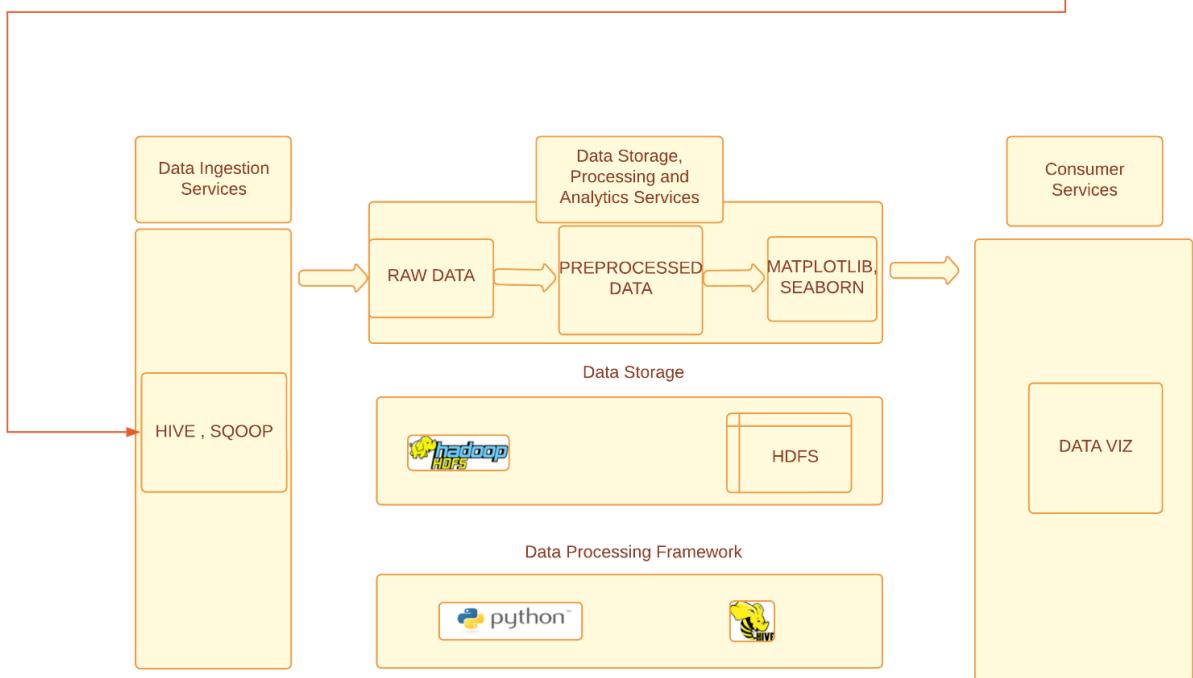
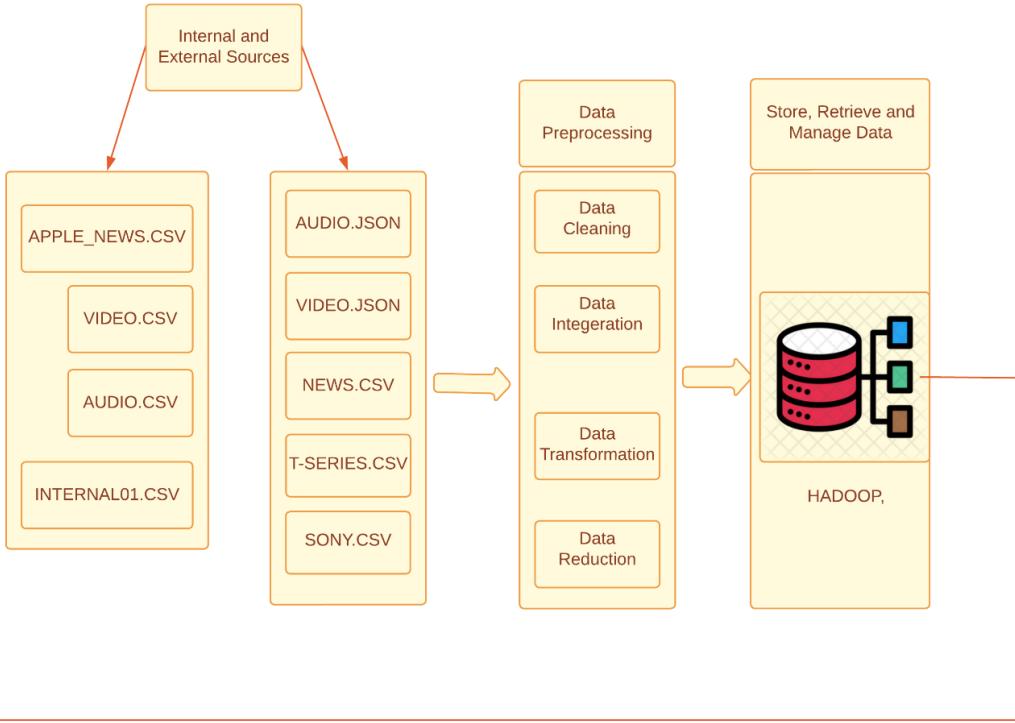
2. The goal of the project

The goal of the project is to find out the number of Apple device users that use in-built Apple apps and also users that use external apps, with the help of the company's own data as well as data provided by vendors.

3. Project Architecture

1. A Linux file server receives a total of 8 files out of which 5 are from external vendors. The external files are in the form of csv, json, and spreadsheet with count 2, 2, and 1 respectively. The files are coming from the external usage of applications in well established Apple products. Sources are based on methodology or scraping.
2. There are also some traditional databases based on the external app usage which are provided by the vendors (Sony and T-Series) for video and audio streaming data details.
3. The file data and tables are validated, enriched, and processed before loading into HDFS/HIVE.
4. After analysing queries and using the visualisation tools, we test the result and use it for enhancing the Apple product data usage of external and internal apps.

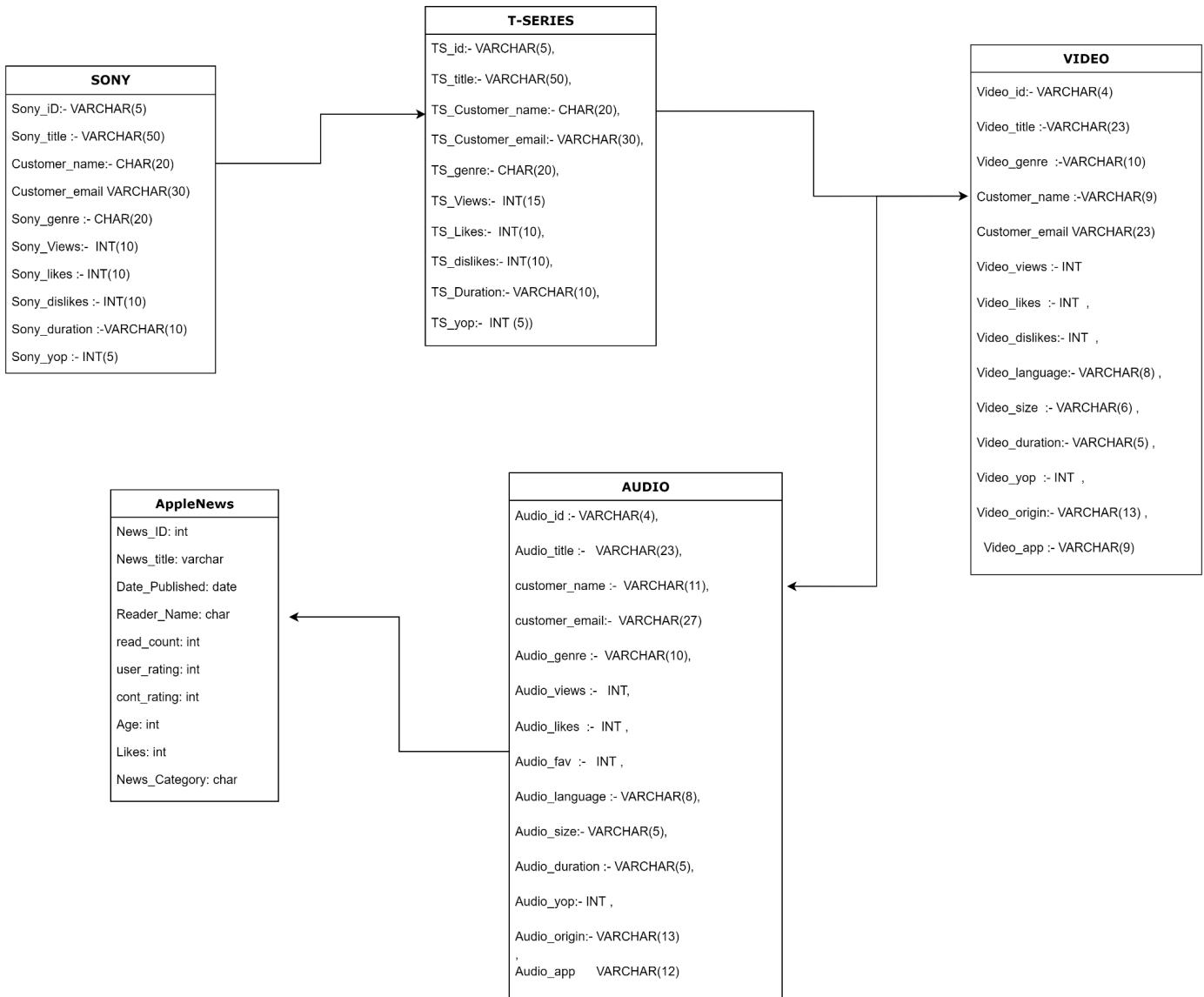
APPLE PRODUCT AND APP USAGE ARCHITECTURE



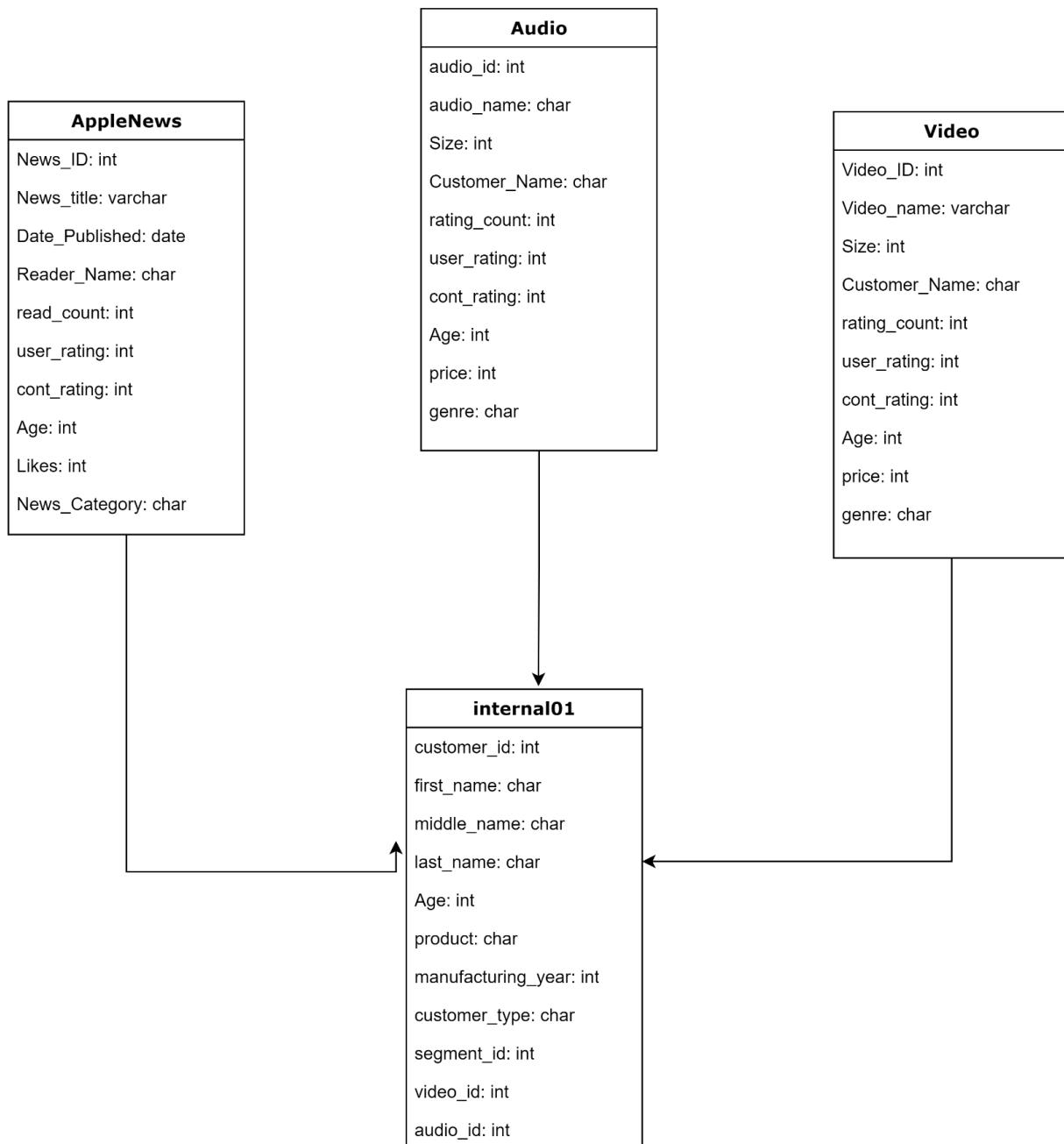
4. Dataset Explanation and Schema

1. Data coming from third-party sources reside in the local directory and has a csv, a json format and a Spreadsheet.
2. A master table is created from the existing tables which is then inserted in HDFS using Sqoop import.
3. Further the Dataset is Visualised using the python Libraries. Pyspark tool is used to map the dataset.

EXTERNAL SCHEMA



INTERNALSCHEMA



Fields present in the data files and tables- External App Usage

Data files contain below fields

Column Name/Field Name Column Description/Field Description

Sony Music Vendor

CSV File Fields

- **Sony_iD** Unique identifier for music
- **Customer_Name** Name of the customer
- **Customer_email** Email of the customer
- **Sony_genre** Genre of the music
- **Sony_views** Numbers of views the music secured
- **Sony_likes** Number of likes the music secured
- **Sony_dislikes** Number of dislikes the music secured
- **Sony_duration** Length of the music
- **Sony_yop** Year the music was published

TSeries Vendor

Another CSV File Fields

- **Ts_iD** Unique identifier for music
- **Ts_title** Title of the music
- **Ts_Customer_Name** Name of the customer
- **Ts_Customer_Email** Email of the customer
- **Ts_genre** Genre of the music
- **Ts_views** Numbers of views the music secured
- **Ts_likes** Number of likes the music secured
- **Ts_dislikes** Number of dislikes the music secured
- **Ts_duration** Length of the music
- **Ts_yop** Year the music was published

Json File Fields

Video Streaming Data

- **Video_id** Unique identifier for Video Streaming
- **Video_title** Title of the Video
- **Customer_Name** Name of the customer
- **Customer_Email** Email of the customer
- **Video_genre** Genre of the Video
- **Video_views** Numbers of views Video secured
- **Video_likes** Number of likes the Video secured
- **Video_dislikes** Numbers of Dislikes Video secured
- **Video_language** Language of the Video

- **Video_size** Size of Video file in MB
- **Video_duration** Length of the Video
- **Video_yop** Year the Video was Published
- **Video_origin** Country that Video belongs to
- **Video_app** The app used to stream Video

Another Json File Fields

Audio Streaming Data

- **Audio_id** Unique identifier for Audio Streaming
- **Customer_name** Name of the customer
- **Customer_email** Email of the customer
- **Audio_fav** Favorite audio
- **Audio_title** Title of the Audio
- **Audio_genre** Genre of the Audio
- **Audio_views** Numbers of views Audio secured
- **Audio_likes** Number of likes the Audio secured
- **Audio_language** Language of the Audio
- **Audio_size** Size of Audio file in MB
- **Audio_duration** Length of the Audio
- **Audio_yop** Year the Audio was Published
- **Audio_origin** Country that Audio belongs to
- **Audio_app** The app used to stream Audio

SpreadSheet File Fields

News Data

- **News_ID** Identifier for news
- **Category** Category articles belongs to
- **Headline** Headline of the article
- **Authors** Person authored that particular article
- **Link** Link to the post

- **Short_description** Short description of the article
- **Date** Date of the article was published
- **Country** Country which news belongs to
- **Language** Regional or national news
- **Channel** News Channel in which news is published

Fields present in the data files and tables- Internal App Usage

Data files contain below fields

Column Name/Field Name Column Description/Field Description

Apple Music App

- **Track_ID** Unique identifier for audio tracks on Apple Music
- **Track_name** Title of the audio track
- **Size** Size of the audio in bytes
- **Customer Name** Name of the customer
- **Rating_Count** No of times the track is rated
- **User_Rating** Rating given by the user to the track
- **Cont_Rating** Rating given by contenders
- **Age** Age of the user
- **Price** Price of the track
- **Genre** Genre of the track

iTunes Video App

- **Video_ID** Unique identifier for video tracks on Apple Music
- **Video_name** Title of the video track
- **Size** Size of the video in bytes
- **Customer_Name** Name of the customer
- **Rating_Count** No of times the video is rated
- **User_Rating** Rating given by the user to the video
- **Cont_Rating** Rating given by the contender to the video
- **Age** Age of the user
- **Price** Price of the video
- **Genre** Genre of the video

Apple News App

- **News_ID** ID of the news
- **News_Title** Title of the news
- **Date_Published** The date on which the news was published
- **Reader_Name** Name of the reader
- **Read_Count** Number of times the news was read
- **User_Rating** Rating given to the news by the user
- **Cont_Rating** Rating given by contenders
- **Age** Age of the reader
- **Likes** Number of likes on the news
- **News_Category** Category of the news.

5. Problem Statement/Tasks

1. Find out the number of users enrolled in Apple devices like iPhone,iPad, Macbook, Apple desktop etc. and use Apple apps.
2. Find out which app is used by the user for listening to music, watching videos, and reading news.
3. Divide the customers using video services according to age group - under 18 and above 18.
4. Find out the average age of users using each Apple product.
5. Popularity of Apple apps vs other apps for users using Apple devices.
6. To find out the popularity of music labels.

6. Code Templates

6.1 Data Processing

6.1.1 Conversion of raw data to processed data:

For each raw file we have checked null values, duplicate values and other parameters and then converted them into processed dataset. Here are some samples of codes.

The screenshot shows a Jupyter Notebook interface with the title "jupyter Indian customers Last Checkpoint: 4 hours ago (autosaved)". The toolbar includes File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Trusted, Python 3 (ipykernel), and Logout. Below the toolbar, there are buttons for Run, Cell, Code, and Cell Type. The notebook has three cells:

- In [1]:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
```
- In [2]:

```
1 df=pd.read_csv('newmaster.csv')
```
- In [3]:

```
1 df.head()
```

Out[3]:

	Customer_Id	First_Name	Last_Name	Middle_Name	Age	Audio_Id	Audio_Name	Product_Id	Product_Name	Music_App	App_Id	Video_Id	Music_Label
0	CU1301	Mayur	Dolifode	NaN	8	A051	Param Sundari	P1	ipad	Amazone Music	MA1	NaN	Sony Music
1	CU1302	Vaishnav	Kale	Gajanan	9	A052	Mehndi Hai Rachnewali	P2	macbook	Amazone Music	MA1	NaN	Sony Music
2	CU1303	Rutwik	Gormode	Pravin	7	A053	Kaise Mujhe Tum	P3	iphone	Spotify	MA2	NaN	T-Series
3	CU1304	Shantanu	Mangale	Ganesh	8	A054	Hum Nashe Mein Toh Nahin	P2	macbook	Apple Music	MA3	NaN	T-Series
4	CU1305	Tarak	Kalore	Sanjay	7	A055	Rehnuma	P3	iphone	Apple Music	MA3	NaN	T-Series

In [4]:
1 df['Middle_Name'].fillna('No middle name', inplace=True)

In [5]:
1 df.head()

Out[5]:

	Customer_Id	First_Name	Last_Name	Middle_Name	Age	Audio_Id	Audio_Name	Product_Id	Product_Name	Music_App	App_Id	Video_Id	Music_Label
0	CU1301	Mayur	Dolifode	No middle name	8	A051	Param Sundari	P1	ipad	Amazone Music	MA1	NaN	Sony Music
1	CU1302	Vaishnav	Kale	Gajanan	9	A052	Mehndi Hai Rachnewali	P2	macbook	Amazone Music	MA1	NaN	Sony Music
2	CU1303	Rutwik	Gormode	Pravin	7	A053	Kaise Mujhe Tum	P3	iphone	Spotify	MA2	NaN	T-Series
3	CU1304	Shantanu	Mangale	Ganesh	8	A054	Hum Nashe Mein Toh Nahin	P2	macbook	Apple Music	MA3	NaN	T-Series
4	CU1305	Tarak	Kalore	Sanjay	7	A055	Rehnuma	P3	iphone	Apple Music	MA3	NaN	T-Series

```
In [22]: 1 df['Music_App']=df['Music_App'].replace('Spotify','Spotify App')

In [23]: 1 df['Music_App'].value_counts()

Out[23]: Apple music    64
Spotify App      52
Amazon music     33
Name: Music_App, dtype: int64

In [24]: 1 df['Product_Name']=df['Product_Name'].replace('ipad','Ipad')
2 df['Product_Name']=df['Product_Name'].replace('macbook','Macbook')
3 df['Product_Name']=df['Product_Name'].replace('iphone','Iphone')
4

In [25]: 1 df['Product_Name']=df['Product_Name'].replace('Ipad ','Ipad')

In [26]: 1 df['Product_Name']=df['Product_Name'].replace('Macbook ','Macbook')

In [27]: 1 sns.set(style='dark')
2 plt.xticks(rotation=30,horizontalalignment='center')
3 sns.countplot(x=df['Music_App'],hue=df['Product_Name'])
4 plt.title("Popularity of music apps for customer using apple products ")
5 plt.xlabel("Music app")
6 plt.ylabel("Count")
```

6.1.2 Processed Data

Internal

	A	B	C	D	E	F	G	H	I	J
1	audio_id	audio_name	Size	Customer_Name	rating_count	user_rating	cont_rating	Age	price	genre
2	281656475	PAC-MAN Premium	100788224	Rishi Sinha	21292	4 4+		16	3.99	Games
3	281796108	Evernote - stay organ	158578688	Alex K	161065	4 4+		45	0	Productivity
4	281940292	WeatherBug - Local	100524032	Neha Dixit	188583	3.5 4+		48	0	Weather
5	282614216	eBay: Best App to Bu	128512000	Arvind Kumar	262241	4 12+		32	0	Shopping
6	282935706	Bible	92774400	Aakriti Soni	985920	4.5 4+		20	0	Reference
7	283619399	Shanghai Mahjong	10485713	Paras Shamma	8253	4 4+		12	0.99	Games
8	283646709	PayPal - Send and re	227795968	Sameek Khan	119487	4 4+		29	0	Finance
9	284035177	Pandora - Music & R	130242560	Jeffery Thomas	1126879	4 12+		34	0	Music
10	284666222	PCalc - The Best Cal	49250304	Jaspreet Singh	1117	4.5 4+		22	9.99	Utilities
11	284736660	Ms. PAC-MAN	70023168	Alexis Rose	7885	4 4+		26	3.99	Games
12	284791396	Solitaire by Mobility!	49618944	David	76720	4.5 4+		46	4.99	Games
13	284815117	SCRABBLE Premium	227547136	Eugenio Levi	105776	3.5 4+		55	7.99	Games
14	284847138	Bank of America - M	160925696	Jodie Comer	119773	3.5 4+		29	0	Finance
15	284876795	TripAdvisor Hotels Fl	207907840	Troye Sivan	56194	4 4+		48	0	Travel
16	284910350	Yelp - Nearby Restau	167407616	Fletcher	223885	4 12+		24	0	Travel
17	285005463	Cross Bandicoot Nit	10735026	Saman Hasan	31456	4 4+		18	2.99	Games
18	285994151	:) Sudoku +	6169600	Abhay Singh	11447	5 4+		17	2.99	Games
19	293622097	Google Earth	37214208	Ishika Jain	446185	3.5 4+		62	0	Travel
20	295646461	The Weather Channel	199734272	Rohan	495626	3.5 4+		55	0	Weather
21	298596830	NRJ Radio	53426176	Leah Lewis	38	3.5 4+		14	0	Music

iTunes App File

	A	B	C	D	E	F	G	H	I	J
1	audio_id	audio_name	Size	Customer_Name	rating_count	user_rating	cont_rating	Age	price	genre
2	281656475	PAC-MAN Premium	100786224	Rishi Sinha	21292	4 4+		16	3.99	Games
3	281796108	Evernote - stay organ	158576688	Alex K	161065	4 4+		45	0	Productivity
4	281940292	WeatherBug - Local	100524032	Neha Dixit	188583	3.5 4+		48	0	Weather
5	282614216	eBay: Best App to Bu	128512000	Arvind Kumar	262241	4 12+		32	0	Shopping
6	282935706	Bible	92774400	Aakriti Soni	985920	4.5 4+		20	0	Reference
7	283619399	Shanghai Mahjong	10485713	Paras Sharma	8253	4 4+		12	0.99	Games
8	283646709	PayPal - Send and re	227795968	Sameer Khan	119487	4 4+		29	0	Finance
9	284035177	Pandora - Music & R	130242560	Jeffery Thomas	1126879	4 12+		34	0	Music
10	284666222	PCalc - The Best Cal	49250304	Jaspreet Singh	1117	4.5 4+		22	9.99	Utilities
11	284736660	Ms. PAC-MAN	70023168	Alexis Rose	7885	4 4+		26	3.99	Games
12	284791396	Solitaire by Mobility!	49618944	David	76720	4.5 4+		46	4.99	Games
13	284815117	SCRABBLE Premium	227547136	Eugene Levi	105776	3.5 4+		55	7.99	Games
14	284847138	Bank of America - M	160925696	Jodie Comer	119773	3.5 4+		29	0	Finance
15	284876795	TripAdvisor Hotels Fl	207907840	Troye Sivan	56194	4 4+		48	0	Travel
16	284910350	Yelp - Nearby Restau	167407616	Fletcher	223885	4 12+		24	0	Travel
17	285005463	Crash Bandicoot Nitr	10735026	Saman Hasan	31456	4 4+		18	2.99	Games
18	285994151	:) Sudoku +	6169600	Abhay Singh	11447	5 4+		17	2.99	Games
19	293622097	Google Earth	37214208	Ishika Jain	46185	3.5 4+		62	0	Travel
20	295646461	The Weather Channel	199734272	Rohan	495626	3.5 4+		55	0	Weather
21	298596830	NRJ Radio	53426176	Leah Lewis	38	3.5 4+		14	0	Music

Apple Video

	A	B	C	D	E	F	G	H	I	J
1	News_ID	News_title	Date_Published	Reader_Name	read_count	user_rating	cont_rating	Age	Likes	News_Category
2	181656475	Stock News	04-04-2022	Rishi Sinha	41292	4 4+		26	399	Stock Market
3	181796108	Pollution in Delhi	04-04-2022	Alex K	11065	4 4+		55	0	City
4	181940292	Weather	04-04-2022	Neha Dixit	188583	3.5 4+		48	34	Weather
5	182614216	Best App to Buy, Sel	04-04-2022	Arvind Kumar	22241	4 12+		52	0	Shopping
6	182935706	Best Places to Visit i	05-04-2022	Aakriti Soni	85920	4.5 4+		20	10	Reference
7	183619399	Crime Rate	05-04-2022	Paras Sharma	253	4 4+		52	99	Crime
8	183646709	PayPal Fraud	06-04-2022	Sameer Khan	1187	4 4+		69	0	Finance
9	184035177	New trending music	07-04-2022	Jeffery Thomas	112679	4 12+		44	230	Music
10	184666222	Forecast for today	08-04-2022	Jaspreet Singh	117	4.5 4+		62	999	Utilities
11	184736660	IPL Updates	09-04-2022	Alexis Rose	785	4 4+		36	399	Sports
12	184791396	T20 News	10-04-2022	David	6720	4.5 4+		46	499	Sports
13	184815117	Sports Highlights	11-04-2022	Eugene Levi	105776	3.5 4+		55	799	Sports
14	184847138	Scam Alert	12-04-2022	Jodie Comer	119773	3.5 4+		39	2	Finance
15	184876795	Most Followed on Inst	13-04-2022	Troye Sivan	6194	4 4+		28	20	Social Media
16	183239304	Top 50 Songs right n	14-04-2022	Fletcher	238885	4 12+		24	120	Travel
17	173842902	Recipe for Summer I	15-04-2022	Saman Hasan	3156	4 4+		58	299	Food
18	185994151	Four Thieves Caught	16-04-2022	Abhay Singh	11447	5 4+		27	229	Crime
19	193622097	Covid Cases Update	17-04-2022	Ishika Jain	46185	3.5 4+		62	10	Health
20	195646461	Unfortunate Events I	18-04-2022	Rohan	495626	3.5 4+		55	111	Crime
21	198596830	New phones under 2	19-04-2022	Leah Lewis	3333	3.5 4+		22	30	Tech

Apple News App

External

	A	B	C	D	E	F	G	H	I	J
1	TS_Id	TS_title	TS_Customer_name	TS_Customer_email	TS_genre	TS_views	TS_likes	TS_dislikes	TS_duration	TS_yop
2	T001	Dhokha	Sneha Tiwari	sneha05@gmail.com	Romance	100000000	1400000	108000	4M53S	2022
3	T002	Barbadnyaan	Ravi Singh	singhrav00@gmail.com	Pop	62008657	6600000	4579	4M02S	2021
4	T003	Vaaste	Raghu Talwar	raghutalwar@gmail.com	Romance	140000000	12000000	348000	4M26S	2019
5	T004	Ik bagal song	Shreyash sinha	shreyassinha098@gmail.com	Indian classical	2205630	42000	2740	5M49S	2021
6	T005	Aankhon Aankhon	Utpal Verma	vermautpal@gmail.com	Disco	7700000	629000	10807	3M49S	2015
7	T006	High rated gabru	Shayam Mukherjee	shyammukherjee@gmail.com	Dance pop	110000000	5700000	146450	3M36S	2017
8	T007	Pehli Baar	Ramya Sachdeva	sachdevaramya@gmail.com	Romantic	38000	1200	26	3M45S	2022
9	T008	Ab Mujhe Raat Din	Kisan Kumar	kisanumar@gmail.com	Romantic	1734809	128000	13760	4M23S	2011
10	T009	One bottle down	Meenakshi roy	meenakshiroy@gmail.com	Jazz	101900000	610000	12845	3M16S	2015
11	T010	Dope Shoppe	Keshav Singh	keshavsingh085@gmail.com	Jazz	158283848	1200000	56893	3M11S	2011
12	T011	Shiddat Title Track	Karunika Goyal	goyalkarunika@gmail.com	Romance	39737533	596000	1678	4M03S	2021
13	T012	GF BF	Shubhankar Sharma	sharmashubhankar@gmail.com	Pop	268098179	1700000	34785	7M02S	2016
14	T013	Lahore	Ram Saxena	saxenaram678@gmail.com	Jazz	1019468215	5500000	44256	3M54S	2017
15	T014	Chhor Denge	Mukesh Sahani	mukeshsahani65@gmail.com	Romantic	421411582	4100000	78078	4m28s	2021
16	T015	Jiya Tu Bihar Ke Lal	Harli Choudhary	choudhary6538@gmail.com	Jazz	21074855	231000	1760	3M23S	2012
17	T016	Nayan	Sudhir Kapoor	sudhirkapoor645@gmail.com	Classical	282435407	3100000	70054	5M59S	2020
18	T017	Agohra	Rakesh Sharma	sharma.rakesh@gmail.com	Jazz	12000000	36000	1860	2M34S	2022

TSERIES

	A	B	C	D	E	F	G	H	I	J
1	sony_id	sony_title	customer_name	customer_email	sony_genre	sonyViews	sony_Likes	sony_Dislikes	sony_Duration	sony_yop
2	s001	Kya Beat Ay	ajit singh	ajitsingh@gmail.com	pop	620,878,106	4100000	4450	3M10S	2018
3	s002	Raataan Lambiyan	ram gopal	gopalram12@gmail.com	pop	666,725,437	6500000	3286	3M13S	2021
4	s003	Genda Phool	priyanka tripathi	ptripathi@gmail.com	pop	900,377,821	6200000	12000	2M53S	2021
5	s004	Naah	arjun singh	singharjun@gmail.com	pop	785677341	5600000	7800	3M22S	2018
6	s005	dynamite	meeta banerjee	mmbanerjee@gmail.com	disco	1,807,938	10000	426	3M39S	2016
7	s006	she's on fire	lewis andrew	andreviewin@gmail.com	dance-pop	33000	281	14	2M45S	2022
8	s007	maahi ve	david peter	peterdavid@gmail.com	rhythm	125,600,274	681000	2120	6M11S	2014
9	s008	dance masti	sunil chouksey	choukseysunil@gmail.com	jazz	26,110,138	109000	9887	5M31S	2015
10	s009	kmamat Se	manjari shri	shrimanjar@gmail.com	pop	676,815	5000	128	4M52S	2014
11	s010	mea da ladla	smith johnson	johnsonsmith@gmail.com	synthpop	670027653	140000	5643	2M46S	2014
12	s011	tunak tunak tun	willy williams	williamawilly@gmail.com	bhangra music	193,804,213	2600000	1400	4M16S	2014
13	s012	mausam	shubhi sharma	sharmashubhi@gmail.com	rhythm and blues	210,150	536	94	3M06S	2012
14	s013	tum aaye	ramya pandit	panditramya@gmail.com	rhythm	109,883	1300	142	4M33S	2014
15	s014	barfi (title track)	martin lopez	lopezmartin@gmail.com	pop	3,160,031	14477	780	4m34s	2012
16	s015	bhaag milka bhaag	jack wilson	wilsonjaci@gmail.com	rock	17,288,866	176000	1780	2M38S	2013
17	s016	raanjhanaa	gaurev rajput	rajputgaurev@gmail.com	classical	20,203,235	240000	2700	4M08S	2013
18	s017	tum tak	martin taylor	taylormartin@gmail.com	pop	26,748,286	202000	1800	5M03S	2013
19	s018	renjha	arpit peter	peterarpit@gmail.com	pop	252,708,525	3040000	2200	3M13S	2021
20	s019	kabhi khushi kabhi g	raj awasthi	awasthira@gmail.com	popular music	58,506,600	353,000	1346	5M13S	2014
21	s020	lakshya (title track)	harry thompson	thompsonharry@gmail.com	rock	43,016,960	397000	3400	4M41S	2015

SONY

	A	B	C	D	E	F	G	H	I	J	K	L
1	New_Id	category	headline	authors	link	short_description	date	Country	Language	Channel	Customer Name	Customer mail id
2	N001	EDUCATION	NTA Declares GATE A Mariyam alavl	http://www.indiatoda	Educational update	23-05-2022	INDIA	English	INDIA TODAY	Ashutosh Mukherji	Ashutosh.Mukherji@gmail.com	
3	N002	SPORTS	patidar from unisold & AK Sikri	http://www.aajtak.in	sports details	24-05-2022	INDIA	English	AAJ TAK	Dr. Salim	Salim223@gmail.co	
4	N003	GOVERNMENT	modi praised tMIL IN Aabshar hassan	http://www.aajtak.in	politics updates	24-05-2022	INDIA	English	AAJ TAK	Dr. Srileshna yadav	yadav321@gmail.co	
5	N004	FINANCIAL	A mutual Fund for ex A Mariyam alavl	http://www.indiatoda	economics updates	23-05-2022	INDIA	English	INDIA TODAY	Kapil Dev	kapil900@gmail.com	
6	N005	SPORTS	one in the million & AK Sikri	http://www.aajtak.in	sports details	24-05-2022	US	English	AAJ TAK	Dr. Salim Ali	Salim.al3343@gmail.com	
7	N006	EDUCATION	wbjee 2022 response Aabshar hassan	http://www.indiatoda	news plans	23-05-2022	INDIA	HINDI	INDIA TODAY	Vallabhal Patel	patel7777@gmail.co	
8	N007	SPORTS	jyoti yamaji smashes AK Sikri	http://www.aajtak.in	sports details	23-05-2022	INDIA	HINDI	AAJ TAK	Ashutosh Vithwakam	vithwa3343@gmail.co	
9	N008	HEALTH	vaccines offer mode Aarish chhabra	http://www.aajtak.in	health related	23-05-2022	INDIA	HINDI	AAJ TAK	Abdul Ghaffar Khan	khan543@gmail.com	
10	N009	SPORTS	IWL odisha maul me Aabshar hassan	http://www.indiatoda	IPL related	23-05-2022	INDIA	English	INDIA TODAY	Dr. Srileshna Singh	singh0000@gmail.co	
11	N010	GOVERNMENT	In uk, nahuil snippets A Mariyam alavl	http://www.aajtak.in	politics updates	23-05-2022	INDIA	HINDI	AAJ TAK	Nagarjuna	janaaa334@gmail.co	
12	N011	SPORTS	real madrid remain e Aarish sharma	http://www.indiatoda	sports details	23-05-2022	INDIA	HINDI	INDIA TODAY	Anucolel singh	yadav345@gmail.co	
13	N012	HEALTH	8 easy ways to reduc AK Sikri	http://www.aajtak.in	news plans	24-05-2022	US	English	AAJ TAK	Dr. Srileshna patel	srileshna.patel9867@gmail.com	
14	N013	EDUCATION	ras 2021 in kerala d Aarish chhabra	http://www.indiatoda	Educational update	23-05-2022	INDIA	HINDI	INDIA TODAY	Dr. Anurag Narayan	anurag5565@gmail.com	
15	N014	GOVERNMENT	JP Nedda to meet m A Mariyam alavl	http://www.aajtak.in	politics updates	23-05-2022	US	English	AAJ TAK	C R Das	dashti@gmail.com	
16	N015	SPORTS	patidar's brilliant ton Aarish chhabra	http://www.aajtak.in	sports details	23-05-2022	INDIA	HINDI	INDIA TODAY	Sallim Ali	ali.khan345@gmail.co	
17	N016	FINANCIAL	RBI govt transfer fail Aabshar hassan	http://www.indiatoda	economics updates	23-05-2022	INDIA	HINDI	INDIA TODAY	Vallabhal Patel	patelvallabhal@gmail.com	
18	N017	EDUCATION	NTA Declares GATE A Mariyam alavl	http://www.indiatoda	news plans	23-05-2022	US	HINDI	AAJ TAK	Chitta Ranjan Das	ranjan7777@gmail.co	
19	N018	SPORTS	united banking on th AK Sikri	http://www.aajtak.in	sports details	23-05-2022	INDIA	HINDI	INDIA TODAY	Miliha Singh	miliha@gmail.com	
20	N019	FINANCIAL	FDI inflow hits all time A Mariyam alavl	http://www.indiatoda	economics updates	23-05-2022		English	AAJ TAK	Kapil Yadav	kapil9002@gmail.co	

NEWS

6.2 Hive and Sqoop

We have used Sqoop to import the data from SQL to Hive and there we can perform our necessary tasks to get the outputs.

Here is the APPLE Database created in Hive.

```
Ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
CU1818 Anil Joseph V 17 A006 Heer Jaan P3 iPhone Spotify MA2
CU1819 Rohith Sathvika S 17 A022 Ipad Rock
CU1820 Jay Sathvika 16 A011 Rock N Roll P3 iPhone Apple music MA3
CU1821 Kiran Rao 13 A021 How It Breaks P3 iPhone Apple music MA3
CU1822 Kavya Mohan S 17 A008 Saare Bolo Bewafa P1 Ipad Apple music MA3
CU1823 Karan Deep 14 A007 Pasoori P3 iPhone Apple music MA3
CU1824 ParvathiL Arun R 16 A014 Po Ve Po P3 iPhone Apple music MA3
CU1825 Pranoy Sathyia M 17 A023 Reset P3 iPhone Apple music MA3
CU1826 Deepthi K 18 A007 Pasoori Macbook Apple music MA3
CU1827 Arun David L 17 A016 Woh Thik Kolaveri Ol P3 iPhone Apple music MA3
CU1828 Anir Navaz K 16 A015 The Rhythms Of Love P3 iPhone Apple music MA3
CU1829 Abhilash Jacob 17 A010 Ranjha P3 iPhone Apple music MA3
CU1830 Jaison Mathew V 17 A028 Beautiful P1 Ipad Apple music MA3
CU1831 Theertha Ramesh 15 A002 Dhokha P1 Ipad Apple music MA3
CU1832 Tarun Rajeev 17 A002 Dhokha P1 Ipad Apple music MA3
CU1833 Adithya L 17 A027 Hot Creek Is Your Love Macbook Apple music MA3
CU1834 Tanya Abraham K 17 A009 Mehacobba P3 iPhone Spotify MA2
CU1835 Rahul Subramanian G 17 A006 Meri Jaan P3 iPhone Spotify MA2
CU1836 Arun Neelakandan 17 A006 Meri Jaan P3 iPhone Spotify MA2
CU1837 Joseph Alex 17 A009 Mehacobba P3 iPhone Spotify MA2
CU1838 Rose John Elizabeth 16 A006 Meri Jaan P3 iPhone Spotify MA2
CU1839 Anitha Arith 16 A030 Perfect P3 iPhone Apple music MA3
CU1840 Akila Francis Paul 17 A026 Good Day P2 Macbook Apple music MA3
CU1841 Diya Das Sanjay 14 A025 The Boys P1 Ipad Spotify MA2
CU1842 Nivaya Samuel John 17 A002 Dhokha Apple music MA3
CU1843 Arthana B Aravind 17 A009 Mehacobba Apple music MA3
CU1844 Arjun Ram 17 A009 Mehacobba Spotify MA2
CU1845 Pranav Sundar T 16 A0031 Excuses
CU1846 Hitesh Ramesh 17 A002 Pasoori Spotify MA2
CU1847 Nitthya Mahesh 17 A002 Dhokha T-Series
CU1848 Vilney Vilney V 15 A007 Pasoori Apple music MA3
CU1849 Ananya Tom Merlin 15 A024 Gangnam Style
CU1850 Keerthi Raveendran R 15 A029 My Love Spotify MA2
CU1851 Rakesh Ravi 17 A0111 Darsana P1 Ipad Spotify MA2
CU1852 Vinweeth Sharma T 17 A0113 Puthiyorum Lokam P1 Ipad Spotify MA2
Time taken: 2.711 seconds, Fetched: 52 row(s)
hive (AppleProject)> select * from under_18 limit 10 ;
OK
Customer_Id First_Name Last_Name Middle_Name NULL Audio_Id Audio_Name Product_Id Product_Name Music_App App_Id Video_Id
uslc_Label
CU1801 Akash Praveen M 17 A001 Heer Raanjhana P1 Ipad Spotify MA2 Sony Music
CU1802 Anal George 16 A002 Dhokha P2 Macbook Spotify MA2 T-Series
CU1803 Aryadeep Chaudhary 17 A001 Ipad Rock Mehacobba P2 Macbook Apple music MA3
CU1804 Arvind Mesare 16 A013 Meri Zindagi Hel Tu P2 Macbook Spotify MA2
CU1805 Sanjeet Dhok 17 A007 Pasoori P2 Macbook Spotify MA2
CU1806 Santosh Makhlia Anyea 17 A004 Kuch Baateln P3 iPhone Apple music MA3
CU1807 Nillesh Tiwari Suresh 16 A004 Kuch Baateln P1 Ipad Apple music MA3
CU1808 Rajendra Mittal Ramesh 17 A030 Perfect P1 Ipad Apple music MA3
CU1809 Prakash Munger Ram 16 A005 Do Mujhe Deewana Kar De P2 Macbook Amazon music MA1
Time taken: 0.425 seconds, Fetched: 10 row(s)
hive (AppleProject)>
```

```

Ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Sun 15:07 ~ ubh01@ubh01:~ 
File Edit View Search Terminal Help
> now format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 0.998 seconds
hive (appleproject)> show tables;
OK
under13
Time taken: 0.047 seconds, Fetched: 1 row(s)
hive (appleproject)> select * from under13;
OK
Customer_Id First_Name Last_Name Middle_Name Age Audio_IdAudio_Name Product_Id Product_Name Music_App App_Id Video_IdMusic_Label
CU1301 Pankaj Kumar Dofode 8 A062 Abhishek Taran Singh P1 Ipad Amazon Music MA1 Sony Music
CU1302 Vaishali Kale Gajanan 9 A052 Hemant Hal Bachewall P2 macbook Amazone Music MA1 Sony Music
CU1303 Rutwik Gornode Pravin 7 A053 Kaise Mujhe Tun P3 Iphone Spotify MA2 T-Series
CU1304 Shantanu Mangale Ganesh 8 A054 Hum Nashe Mein Toh Nahin P3 macbook Apple Music MA3 T-Series
CU1305 Tarak Kalore Sanjay 7 A055 Rehnuma P3 iphone Apple Music MA3 T-Series
CU1306 Sandeep Tawate Pradeep 5 A056 Kun Faya Kun P3 Iphone Apple Music MA3 Sony Music
CU1307 Sambhu Bhambhani Dharmale Gurav 6 A057 Meri Zara Si P1 Ipad iPhone Apple Music MA3 Sony Music
CU1308 Vidhi Verna Galanjan 12 A058 Chakka Chakkathi P1 Ipad Amazon Music MA1 Sony Music
CU1309 Charlie Munger Galanjan 10 A059 Kaalame P1 Ipad Amazon Music MA1 Sony Music
CU1310 Narendra Kushwaha 12 A060 Ishq Bezbahan P2 macbook Amazone Music MA1 T-Series
CU1311 Sunit Shrivastava kunar 8 A061 Jalwanuma P2 macbook Spotify MA2 T-Series
CU1312 Neeraj Mishra 6 A062 Komuram Bheemudo P3 Iphone Spotify MA2 T-Series
CU1313 Rakesh Saini 11 A063 Anupam Kaikai P1 Ipad Spotify MA2 Sony Music
CU1314 Gagan Singhpoor Singh 10 A064 Mai Tujhe Salaam P1 Ipad Apple Music MA3 Sony Music
CU1315 Pratihvi Pandey 9 A065 The Human Song P1 Ipad Apple Music MA3 T-Series
CU1316 Akshita Jain 5 A066 Mast Nazron se P3 Iphone Apple Music MA3 T-Series
CU1317 Nishtha Mehta 8 A067 Teri Ankho Mein P1 Ipad Apple Music MA3 T-Series
CU1318 Bhavna Yadav 7 A068 Pathhaar Wargi P3 Iphone Amazon Music MA1 T-Series
CU1319 Pankaj Valswakarna 9 A069 High Raah P2 Ipad Spotify MA2 T-Series
CU1320 Puneet Vishwakarma 6 A070 Genda Phool P2 macbook Spotify MA2 Sony Music
CU1321 Sahitya Gour Kumar 9 A071 Backbone P3 Iphone Apple Music MA3 T-Series
CU1322 Ankit Chouhan Singh 10 A072 Soch P2 macbook SpotifyMA2 Sony Music
CU1323 Pratyush Tiwari Kumar 11 A073 Soch na Sahe-Sahe Tera P2 macbook Amazone Music MA1 Sony Music
CU1324 Aayush Chouhan Singh 12 A074 Tiltlaa Marga P1 Ipad Amazon Music MA1 Sony Music
CU1325 Priti Sahu 8 A075 Aashiq Banaya Apne P1 Ipad Apple Music MA3 T-Series
CU1326 Nitesh Chourasiya 7 A076 Vaaste macbookApple Music MA3 T-Series
CU1327 Gourav Guru 12 A077 Naam P3 Iphone Apple Music MA3 Sony Music
CU1328 Dilgyljyot Singh 11 A078 Tunse Pyar Karke P1 Ipad Apple Music MA3 Sony Music
CU1329 Neetu Singh 12 A079 Lut Gaye P1 Ipad Spotify MA2 T-Series
CU1330 Lovely Gupta 5 A080 Loca P3 Iphone SpotifyMA2 T-Series
CU1331 Ashish Kumar Gupta 6 A081 Amlit Narleen Kahbi Rehde P3 Iphone Amazone Music MA1 T-Series
CU1332 Harpreet Choudhary Kumar 8 A082 Tujhe Bhul Diya P2 macbook Amazone Music MA1 T-Series
CU1333 Santosh Sahu Kumar 7 A083 Jee karr Daa P3 Iphone Apple Music MA3 Sony Music
CU1334 Jitendra Dubey kumar 6 A084 Jokar P2 macbookAmazone Music MA3 Sony Music
CU1335 Kuldeep Vlshwakarma 8 A085 Kabhi yun P2 macbook Apple Music MA3 T-Series
CU1336 Gaurav Deepwan 11 A086 Kuch Baatein P2 macbonkAmazone Music MA1 Sony MUSIC
Time taken: 2.072 seconds, Fetched: 52 rows(s)
hive (appleproject)> create table under18 customer_id string, first_name string, last_name string, middle_name string,
> age string, audio_id string, audio_name string, product_id string, product_name string,
> music_app string, app_id string, video_id string, music_label string)
> row format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 0.191 seconds
hive (appleproject)> create table under18(customer_id string, first_name string, last_name string, middle_name string,
> age string, audio_id string, audio_name string, product_id string, product_name string,
> music_app string, app_id string, video_id string, music_label string)
> row format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 0.185 seconds
hive (appleproject)> select * from under18 limit 10;
OK
Time taken: 0.256 seconds
hive (appleproject)> select * from under18 limit 10;
OK
Customer_Id First_Name Last_Name Middle_Name Age Audio_IdAudio_Name Product_Id Product_Name Music_App App_Id Video_IdMusic_Label
CU1381 Akash Praveen M 17 A001 Heer Raajhana P1 Ipad Spotify MA2 Sony Music
CU1382 Anal George 16 A002 Dhokha P2 Macbook Spotify MA2 T-Series
CU1383 Aaryadeep Chaudhary Praveen 17 A009 Mehabooba P2 Macbook Apple music MA3 T-Series
CU1384 Shreyash Mesare 16 A013 Meri Zindagi Hai Tu P2 Macbook Spotify MA2 T-Series
CU1385 Sanjeet Dhok 17 A007 Pasoor P2 Macbook Spotify MA2 Cool Studio T-Series
CU1386 Kunal Mukhipa Ameys 17 A004 Kuch Baatain P2 Iphone Apple music MA3 T-Series
CU1387 Nilesh Tiwari Suresh 16 A004 Kuch Baattein P1 Ipad Apple music MA3 T-Series
CU1388 Rajendra Mittal Ramesh 17 A030 Perfect P1 Ipad Apple music MA3 T-Series
CU1389 Prakash Munger Ram 16 A005 Jo Mujhe Deewana Kar DeP2 Macbook Amazon music MA1 T-Series
Time taken: 0.266 seconds, Fetched: 10 row(s)
hive (appleproject)> 

```

```

Ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Sun 15:07 ~ ubh01@ubh01:~ 
File Edit View Search Terminal Help
> now format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 0.998 seconds
hive (appleproject)> show tables;
OK
under18
Time taken: 0.047 seconds, Fetched: 1 row(s)
hive (appleproject)> select * from under18;
OK
Customer_Id First_Name Last_Name Middle_Name Age Audio_IdAudio_Name Product_Id Product_Name Music_App App_Id Video_IdMusic_Label
CU1381 Akash Praveen M 17 A001 Heer Raajhana P1 Ipad Spotify MA2 Sony Music
CU1382 Anal George 16 A002 Dhokha P2 Macbook Spotify MA2 T-Series
CU1383 Aaryadeep Chaudhary Praveen 17 A009 Mehabooba P2 Macbook Apple music MA3 T-Series
CU1384 Shreyash Mesare 16 A013 Meri Zindagi Hai Tu P2 Macbook Spotify MA2 T-Series
CU1385 Sanjeet Dhok 17 A007 Pasoor P2 Macbook Spotify MA2 Cool Studio T-Series
CU1386 Kunal Mukhipa Ameys 17 A004 Kuch Baatain P2 Iphone Apple music MA3 T-Series
CU1387 Nilesh Tiwari Suresh 16 A004 Kuch Baattein P1 Ipad Apple music MA3 T-Series
CU1388 Rajendra Mittal Ramesh 17 A030 Perfect P1 Ipad Apple music MA3 T-Series
CU1389 Prakash Munger Ram 16 A005 Jo Mujhe Deewana Kar DeP2 Macbook Amazon music MA1 T-Series
Time taken: 0.266 seconds, Fetched: 10 row(s)
hive (appleproject)> 

```

Apple Products and App Usages

```

Ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
[  ] CU1300 Mayur Poltode 8 A051 Param Sundar P1 Ipad Amazon Music MA1 Sony Music
CU1302 Krishnav Kale Gajanan 9 A052 Kavindri Hira Rachmali P2 Macbook Amazon music MA1 Sony Music
CU1303 Rutvik Gormade Pravin 7 A053 Kaise Mushe Tum P3 Iphone Spotify MA2 T-Series
CU1304 Shantanu Mangale Ganesh 8 A054 Hum Nashe Mein Toh NahinP2 Macbook Apple Music MA3 T-Series
CU1306 Sandeep Tatewar Pradeep 5 A055 Rehnuma P3 Iphone Apple Music MA3 T-Series
CU1307 Pradumn Dharmale Gunwant A057 Kun Faya Kun P3 Iphone Apple Music MA3 Sony Music
CU1308 Patel Verma Gajjanan 10 A058 Chakka Chalak P1 Ipad Amazon Music MA1 Sony Music
CU1309 Charlie Hunger 12 A059 Kaalame P1 Ipad Amazon Music MA1 Sony Music
CU1310 Narendra Kushwaha 12 A060 Ishq Bezaaban P2 Macbook Spotify MA2 T-Series
CU1311 Sumit Shrivastava Kumar 8 A061 Jalwanuma P2 Macbook Spotify MA2 T-Series
CU1312 Neeraj Mishra 6 A062 Konram Bheemudo P3 Iphone Spotify MA2 T-Series
CU1313 Ayush Swarn 11 A063 Masse P1 Ipad Spotify MA2 Sony Music
CU1314 Arunjeet Singh 10 A064 Main Tu Jhene Salam P3 Ipad Apple Music MA3 Sony Music
CU1315 Prathyl Pandey 9 A065 The Huma Song P1 Ipad Apple Music MA3 T-Series
CU1316 Akshita Jain 5 A066 Mast Nazron se P3 Iphone Apple Music MA3 T-Series
CU1317 Nishtha Mehta 8 A067 Terti Ankho Meln P1 Ipad Apple Music MA3 T-Series
CU1318 Bhavna Yadav 7 A068 Pathkar Wargi P3 Iphone Amazon Music MA1 T-Series
CU1319 Kapil Vishwakarma 6 A069 High Rated P3 Iphone Spotify MA2 T-Series
Time taken: 0.17 seconds
Fetched: 20 row(s)
hive (appleproject)> create table sony_data(song_id string, sony_title string, customer_name string, customer_email string, sony_genre string, sony_views string, sony_likes string, sony_dislikes string, sony_duration string, sony_yop string)
> row format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 0.149 seconds
hive (appleproject)> select * from sony_data limit 20;
OK
sony_id sony_title customer_name customer_email sony_genre sony.Views sony_Likes sony_DisLikes sony_duration sony_yop
s001 Kya Baat Ay ajit singh ajitsingh@gmail.com pop "620" 878 106" 4100000 4458
s002 Kya Baat Ay abhishekbitiya@gmail.com pop "655" 973 107" 4327 6280000 3286
s003 Genda Phool priyanka tripathi ptripathi@gmail.com pop "900" 377 82" 6280000 12000
s004 Naah arjun singh singharpunj@gmail.com pop "78567731" 5600000 7800 3M225 2018
s005 dynamite meeta banerjee mbanerjee@gmail.com disco "1" 807 938" 100000 426
s006 she's on fire lewis andrew andrewlewis@gmail.com dance-pop "33000" 281 14 2M45S 2022
s007 maha ve david pete peterdavid@gmail.com rhythm "125" 600 274" 6810000 2120
s008 suno suno gaurav keya gauravkeya@gmail.com jazz "26" 110 138" 1090000 307
s009 kismat Se manjari shri shrimanjaris@gmail.com rock "676" 815" 50000 128 4M52S
s010 maa da ladla smith johnson johnsonsmith@gmail.com synthpop "870827653" 1400000 5643 2M46S 2014
s011 tunak tunak tun willy williams williamswill@gmail.com bhangra music "193" 804 213" 2600000 1400
s012 mausam shubhl sharma sharmashubh@gmail.com rhythm and blues "210" 150" 536 94 3M06S
s013 tum aaye ramya pandit panditramya@gmail.com rhythm "109" 883" 1800 145 4R03S
s014 tum little trade' sunita lopez sunita_lopez1997@gmail.com pop "3" 100 931" 24477 788
s015 baag milka baag jack wilson wilsonjack@gmail.com rock "17" 288 866" 176000 1780
s016 raanjhanaa gaurav rajput rajputgaurav@gmail.com classical "20" 203 235" 2480000 2700
s017 tum tuk martin taylor taylormarting@gmail.com pop "76" 748 286" 202000 1800
s018 ranjha arpit peter peterarpit@gmail.com pop "252" 708 525" 3040000 2200
s019 kabhi khush kabhi ghan raj awasti awasthira@gmail.com popular music "58" 506 600" "353 000"
Time taken: 0.08 seconds
Fetched: 20 rows
hive (appleproject)>

```

```

Ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
[  ] CU081001 Ally Aagaard 17 AU$001 Despacito(Remix) P1 Ipad Spotify MA2 Sony Music
CU081002 Alex Abadi 17 AU$002 Rap God P2 Macbook Amazon music MA1 Sony Music
CU081003 Luis Alvarado 17 AU$003 Bad And Boujee P3 Macbook Apple music MA3 Warner music group
CU081004 Kylla Abbott Elizabeth 16 AU$004 Shape Of You P2 Macbook Spotify MA2 Warner music group
CU081005 Bryce Abbott 15 AU$005 The Hills P1 Ipad Spotify MA2 Warner music group
CU081006 Daniela Abella 15 AU$006 N95 P1 Ipad Spotify MA2 Warner music group
CU081007 Relly Abraham 15 AU$007 We Cry Together P1 Ipad Amazon music MA1 T-Series
CU081008 Michael Acevedo Nicholas 14 AU$008 Auntie Diaries P3 Iphone Amazon music MA1 T-Series
CU081009 Brianna Adams Zoe 14 AU$009 Mother I Sober P3 Iphone Amazon music MA1 Virgin Records
Time taken: 0.175 seconds
Fetched: 10 row(s)
hive (appleproject)> select * from us_under_18 limit 20;
OK
Customer_Id First_Name Last_Name Middle_Name Age Audio_IdAudio_Name Product_Id Product_Name Music_App App_Id Video_IdMusic_Label
CU081010 Sean Allen S I 8 AU$010 Despacito(Renix) P1 Ipad Spotify MA2 Sony Music
CU081011 Miles Allen 9 AU$012 Rap God P2 Macbook Amazon music MA1 Sony Music
CU081012 Peyton Allen 6 AU$013 Wap P2 Macbook Apple music MA3 Warner music group
CU081013 Lindsay Allen Marie 5 AU$014 Shape Of You P2 Macbook Spotify MA2 Warner music group
CU081014 Lucy Allen 10 AU$015 The Hills P1 Ipad Spotify MA2 Warner music group
CU081015 DoMarea' Allen 11 AU$016 Bad And Boujee P3 Macbook Apple music MA3 Warner music group
CU081016 Lucy Allen 11 AU$017 Despacito(remix) P1 Ipad Spotify MA2 Warner music group
CU081017 Quinton Abraham Russell 12 AU$018 We Cry Together P1 Ipad Amazon music MA1 T-Series
CU081018 Billal Abraham 10 AU$019 Auntie Diaries P3 Iphone Amazon music MA1 Warner music group
CU081019 Isatah Abraham 5 AU$020 Mother I Sober P3 Iphone Amazon music MA1 T-Series
CU081020 Lyam Abraham 9 AU$021 God's Plan P3 Iphone Apple music MA3 Sony Music
CU081021 Harold Abraham 8 AU$022 Rocknode P3 Iphone Apple music MA3 Sony Music
CU081022 Isaac Abraham 4 AU$023 Moon N Moon Hot P3 Ipad Spotify MA2 Sony Music
CU081023 Jessica Adams Marie 3 AU$024 Drunk In Love P3 Iphone Spotify MA2 Warner music group
CU081024 Julissa Adams Rose 2 AU$025 United In Grief P1 Ipad Spotify MA2 T-Series
CU081025 liberty Adams Jo 1 AU$026 Father Time P1 Ipad Amazon music MA1 Virgin Records
CU081026 Josh Adams 4 AU$027 Savior P1 Ipad Amazon music MA1 T-Series
CU081027 Kaitlyn Adcock Barrett 6 AU$028 Worldwide Steppers P1 Ipad Amazon music MA1 T-Series
CU081028 Katelyn Adeeb Selaine 9 AU$029 Die Hard P2 Macbook Apple music MA3 T-Series
Time taken: 0.208 seconds, Fetched: 20 row(s)
hive (appleproject)>

```

Apple Products and App Usages

Ubuntu [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Terminal Sun 15:37

ubho1@ubho1: ~

```

Re: Hoify for your platform... using builtin-Java classes where applicable
put: 'us/under_013.csv': No such file or directory
De:ubho1@ubho1:~$ hadoop fs -put US/under_013.csv /user/hive/warehouse/appleproject
D:./ubho1@ubho1:~$ hadoop fs -put US/under_18.csv /user/hive/warehouse/appleproject
D:Muddy/us/under_18
D:ubho1@ubho1:~$ hadoop fs -put US/under_18.csv /user/hive/warehouse/appleproject
Vid: newmaster
Tr: 22/05/22 15:21:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-Java classes where applicable
ubho1@ubho1:~$ hadoop fs -put sony_data.csv /user/hive/warehouse/appleproject.db
Off: /sony_data
22/05/22 15:21:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-Java classes where applicable
put: 'sony_data.csv': No such file or directory
ubho1@ubho1:~$ hadoop fs -put sony_data.csv /user/hive/warehouse/appleproject.db
/sony_data
22/05/22 15:28:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-Java classes where applicable
ubho1@ubho1:~$ hadoop fs -put sony_data.csv /user/hive/warehouse/appleproject.db
Time taken: 0.008 seconds, Fetched: 20 row(s)
ubho1@ubho1:~[ ]

```

ubho1@ubho1: ~

```

File Edit View Search Terminal Help
377 821" 6200000 12000
s7800 3M225 2018
s005 dynamite neeta banerjee mmbanerjee@gmail.com pop 785677341 5600000
938" 10000 420
s006 shie on fire toria andrew andrewlevis@gmail.com dance-pop 33000
s007 mahl ve david peter peterdavid@gmail.com rhythmn "125 600
281 14" 2M45S 2022
s008 dance nasti sunil chouksey choukseyunil@gmail.com jazz "26 110
13" 10000000 9887
s009 kismat 5d shrinjari shri shrinjanjari@gmail.com pop "676 815"
s000 128 4M52S
s010 maa da ladla smith johnson johnsonsmith@gmail.com synthpop 8700276
as53 140000 5643 2M46S 2014
s011 tunad tunak tun willy williams williamswillly@gmail.com bhangra music "193
s012 naujais shubhl sharma sharmashubhi@gmail.com rhythm and blues "210
s013 tun aye rama pandit panditrama@gmail.com rhythm "109 883"
s014 barfl (title track) martin lopez lopezmartin@gmail.com pop "3
1300 142 4M35S
s015 bhag allama bhag jack wilson wilsonjack@gmail.com rock "17
288 866" 170000 1780
s016 ranjhana gaurav rajput rajputgaurav@gmail.com classical "20
203 235" 240000 2700
s017 tun tak martin taylor taylormartin@gmail.com pop "26 748 286"
2028000 1880
s018 amit arpit peter peterarpit@gmail.com pop "252 708 525"
3040000 2200
s019 kabhi khushi kabhi gham raj awasthi awasthiraj@gmail.com popular music
"58 506 600" "353 000"
Time taken: 0.189 seconds, Fetched: 20 row(s)
hive <appleproject>

```

Ubuntu [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Terminal Sun 15:37

ubho1@ubho1: ~

```

CU001302 Miles Allen 9 AU0552 Rap God P2 Macbook Amazone music MA1 Sony Music
CU001303 Peyton Allen 6 AU0553 Wap P2 Macbook Apple music MA3 Sony Music
CU001304 Lindsay Allen Marie 5 AU0554 Shape Of You P2 Macbook Spotify MA2 Virgin Records
CU001305 DeMarea' Allen 10 AU0555 The Hills P1 Ipad Spotify MA2 Virgin Records
CU001306 lucy Allen 11 AU0556 Nya P1 Ipad Spotify MA2 Virgin Records
CU001307 Quinton Abraham Russell 12 AU0557 Cry Together Ipad Virgin Records
CU001308 Billab Abraham 12 AU0558 Auntie Dtarles P3 Iphone Amazon music MA1 T-Series
CU001309 Isalah Abraham 5 AU0559 Mother I Sober P3 Iphone Amazon music MA1 Warner music group
CU001310 Tonit Abraham Marie 6 AU0560 Bad And Boujee P3 Iphone Apple music MA3 sony Music
CU001311 Harold Abraham 8 AU0561 God's Plan P3 Iphone Apple music MA3 sony Music
CU001312 Isack Abraham 4 AU0562 I'm Not Hot P3 Iphone Spotify MA2 Virgin Records
CU001313 Jessica Adams Marie 3 AU0564 Drunk In Love P3 Iphone Spotify MA2 Warner music group
CU001314 Julissa Adams Rose 2 AU0565 United In Grief P1 Ipad Spotify MA2 T-Series
CU001315 liberty Adams Jo 1 AU0566 Father Time P1 Ipad Amazone music MA1 Virgin Records
CU001316 Josh Adams 1 AU0567 Savior P1 Ipad Amazone music MA1 Virgin Records
CU001318 Kayla Adcock Barrett 6 AU0568 Worldwide Steppers P1 Ipad Amazone music MA1 T-Series
CU001319 Kaitlyn Adcock Schatzky 9 AU0569 Die Hard P2 Macbook Apple music MA3 T-Series
Time taken: 0.208 seconds, Fetched: 20 row(s)
hive <appleproject> create table newmaster(customer_id string, first_name string, last_name string, middle_name string, age string, audio_id string, audio_name string, product_id string, product_name string,
> music_app string, app_id string, video_id string, music_label string)
> row format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 0.166 seconds
hive <appleproject> select * from newmaster limit 20;
OK
customer_id First_Name Last_Name Middle_Name Age Audio_IdAudio_Name Product_Id Product_Name Music_App App_Id Video_IdMusic_Label
CU1301 Mayur Dolfode 8 A051 Param Sundar P1 Ipad Amazone Music MA1 Sony Music
CU1302 Vaishnavi Kale Gajanan 9 A052 Mehndi Hal Bachewall P2 macbook Amazone Music MA1 Sony Music
CU1303 Rutwik Gornode Pravin 7 A053 Kaise Mujhe Tun P3 iphone Spotify MA2 T-Series
CU1304 Shantanu Mangale Ganesh 8 A054 Hun Nashe Mein Toh NahinP2 macbook Apple Music MA3 T-Series
CU1305 Tarak Kalore Sanjali 7 A055 Rehnuma P1 Iphone Apple Music MA3 T-Series
CU1306 Deepak Setwar Deep 5 A056 Kun Faya Kun P1 Iphone Apple Music MA3 Sony Music
CU1307 Pradyumn Dharasale Gunwant C A057 Rait Zara Si P3 iphone Apple Music MA3 Sony Music
CU1308 Vldhil Verna Gajanan 12 A058 Chakka Chakalathi P1 Ipad Amazone Music MA1 sony Music
CU1309 Charlie Munger 10 A059 Kaalame P1 ipad Amazon Music MA1 Sony Music
CU1310 Narendra Kushwaha 12 A060 Ishq Bezuaban P2 macbook Amazone Music MA1 T-Series
CU1311 Suniti Shirvastava kunar 8 A061 Jalwauna P2 macbook Spotify MA2 T-Series
CU1312 Anupam Alshara 6 A062 Konkan Moudud P1 Iphone Spotify MA2 T-Series
CU1313 Ayarn Swarn 11 A063 Jessie's Land P1 Ipad Spotify MA2 sony Music
CU1314 Goutam Rajpoot Singh 10 A064 Maa tujhe Salam P1 ipad Apple Music MA3 Sony Music
CU1315 Prathvi Pandey 9 A065 The Human Song P1 ipad Apple Music MA3 Sony Music
CU1316 Akshita Jain 5 A066 Mast Nazron se P3 iphone Apple Music MA3 T-Series
CU1317 Nishtha Mehta 8 A067 Teri Ankho Mein P1 ipad Apple Music MA3 T-Series
CU1318 Neha Yadav 7 M068 Pathkar Wargt P3 Iphone Amazone Music MA1 T-Series
CU1319 Kapil Vishwakarma 9 A069 High Rated P3 Iphone Spotify MA2 T-Series
Time taken: 0.17 seconds, Fetched: 20 row(s)
hive <appleproject>

```

```
Ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help


```

> music_app_id string, app_id string, video_id string, music_label string)
> row format delimited
> fields terminated by ','
> stored as textfile;
NoVtableException[2011]
at org.apache.hadoop.hive ql.parse.HiveParser.createTableStatement(HiveParser.java:6059)
at org.apache.hadoop.hive ql.parse.HiveParser.ddlStatement(HiveParser.java:3808)
at org.apache.hadoop.hive ql.parse.HiveParser.execStatement(HiveParser.java:2382)
at org.apache.hadoop.hive ql.parse.HiveParser.createStatement(HiveParser.java:1333)
at org.apache.hadoop.hive ql.parse.ParseDriver.parse(ParseDriver.java:208)
at org.apache.hadoop.hive ql.parse.ParseDriver.parse(ParseDriver.java:77)
at org.apache.hadoop.hive ql.parse.ParseDriver.parse(ParseDriver.java:70)
at org.apache.hadoop.hive ql.Driver.compile(Driver.java:468)
at org.apache.hadoop.hive ql.Driver.compileInternal(Driver.java:1317)
at org.apache.hadoop.hive ql.Driver.runInternal(Driver.java:1457)
at org.apache.hadoop.hive ql.Driver.runDriver(Driver.java:1237)
at org.apache.hadoop.hive ql.Driver.run(Driver.java:1227)
at org.apache.hadoop.hive.cli.CliDriver.processDDLCmd(CliDriver.java:233)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:184)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:493)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:886)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException
Line 1:10: line 1 did not recognize input near 'string' '' 'first_name' in create table statement
hive (appleproject): create table us_under_018 as select * from us_under_18 limit 10;
hive (appleproject): select * from us_under_018;
OK
Time taken: 0.161 seconds
hive (appleproject): select * from us_under_18 limit 10;
OK
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|Customer_Id|First_Name|Last_Name|Middle_Name|Age|Audio_Id|Audio_Name|Product_Id|Product_Name|Music_App|App_Id|Video_Id|Music_Label|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|CUS001001|Alyssa|Appard|17|AUS001|Shyshka|P1|Ipad|Spotify|MA2|Sony Music|MA1| |
|CUS018002|Alex|Abdul|17|AUS002|Rap God|P2|Macbook|Amazon music|MA3|Sony Music|MA1|
|CUS018003|Lyla|Abbatello|Nicole|17|AUS003|Wap|P2|Macbook|Apple music|MA3|Sony Music|MA1|
|CUS018004|Kyria|Abbott|Elizabeth|16|AUS004|Shape Of You|P2|Macbook|Spotify|MA2|Warner music group|MA1|
|CUS018005|Bryce|Abbott|15|AUS005|The Hills|P1|Ipad|Spotify|MA2|Warner music group|MA1|
|CUS018006|Daniela|Abella|15|AUS006|N95|P1|Ipad|Spotify|MA2|Warner music group|MA1|
|CUS018007|Danielle|Arahman|15|AUS007|We Cry Together|P1|Ipad|Amazon music|MA1|T-Series|MA1|
|CUS018008|Micaela|Accord|Nicholas|14|AUS008|Dance Diaries|P3|Iphone|Amazon music|MA1|T-Series|MA1|
|CUS018009|Breonna|Adam|Zoe|14|AUS009|Mother I Sober|P3|Iphone|Amazon music|MA1|Virgin Records|MA1|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Time taken: 0.175 seconds, Fetched: 10 row(s)
hive (appleproject):
```


```

Customer_Id	First_Name	Last_Name	Middle_Name	Age	Audio_Id	Audio_Name	Product_Id	Product_Name	Music_App	App_Id	Video_Id	H
CU0101	Shantanu	Kuchे	Vlnod	18	A051	Param Sundari	P1	ipad	Amazona Music	MA1	V01	Sony Music
CU0102	Bikhl	Thakare Mukesh	20	20	A052	Mehndi Hat Rachneewal	P2	macbook	Amazona Music	MA1	V02	Sony Music
CU0103	Aneya	Tale Arvind	19	20	A053	Katse Mujhe Tun P3	P3	iphone	Spotify	MA2	V03	T-Series
CU0104	Nelson	Shetkar	Manoj	20	A054	Hum Nashe Meln	P1	macbook	Apple Music	MA3	V04	T-Series
CU0105	Shreyash	Agarwal	19	20	A055	Mehndi Holi	P2	macbook	Amazona Music	MA1	V05	T-Series
CU0106	Rishabhesh	Deshmukh	Gajan	20	A056	Kun Faya K	P1	iphone	Apple Music	MA3	V06	Sony Music
CU0107	Shreejan	Deshmukh	Ramesh	18	A057	Rait Zara S1	P3	iphone	Apple Music	MA1	V07	Sony Music
CU0108	Pranav	Pawar	Lallit	18	A058	Chakka Chakhalati	P1	ipad	Amazona Music	MA1	V08	Sony Music
CU0109	Sangita	Verna Vilay	Khanzode	20	A059	Kaalame P3	P3	iphone	Amazona Music	MA1	V09	Sony Music
CU0110	Shivam	Patil	Shubham	19	A060	Ishq Bezaab	P2	macbook	Amazona Music	MA1	V10	T-Series
CU0111	Santosh	Gates	Sanjay	20	A061	Konuran Bittu	P3	iphone	Spotify	MA2	V11	T-Series
CU0112	Sanjay	Cook	Ranjeet	19	A062	Jessie's Land	P3	iphone	Spotify	MA2	V12	T-Series
CU0113	Gajanan	Murthy	Palash	20	A064	Maai tujhe Salam	P3	iphone	Apple Music	MA3	V13	Sony Music
CU0115	Sanvi	Ambekar	Satyam	19	A065	The Humma Song	P1	ipad	Apple Music	MA3	V15	Sony Music
CU0116	Shital	Advan	Gautam	20	A066	Mast Nazron se	P3	iphone	Apple Music	MA3	V16	T-Series
CU0117	Shreya	Musik	Shreya	20	A067	Tert Anhko Meln	P3	iphone	Apple Music	MA1	V17	T-Series
CU0118	Ratan	Sharma	Naval	19	A068	Pathmar Wari	P3	iphone	Amazona Music	MA1	V18	T-Series
CU0119	Mukesh	Abdul	Dhirlubhai	18	A069	High Rated	P3	iphone	Spotify	MA2	V19	T-Series

Apple Products and App Usages

The tables created in the databases as mentioned in the schema

Browsing HDFS

Thu 22:54 •

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/hive/warehouse/appleproject.db Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxrwxr-x	ubh01	supergroup	0 B	22/5/2022, 3:28:59 pm	0	0 B	newmaster
drwxrwxr-x	ubh01	supergroup	0 B	22/5/2022, 3:36:12 pm	0	0 B	sony_data
drwxrwxr-x	ubh01	supergroup	0 B	22/5/2022, 3:06:33 pm	0	0 B	under13
drwxrwxr-x	ubh01	supergroup	0 B	22/5/2022, 3:13:15 pm	0	0 B	under18
drwxrwxr-x	ubh01	supergroup	0 B	22/5/2022, 3:11:31 pm	0	0 B	under21
drwxrwxr-x	ubh01	supergroup	0 B	22/5/2022, 3:21:51 pm	0	0 B	us_under_013
drwxrwxr-x	ubh01	supergroup	0 B	22/5/2022, 3:22:25 pm	0	0 B	us_under_18

Hadoop, 2015.

AgeGroup - Google Drive

Sun 15:12 •

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/hive/warehouse/appleproject.db Go!

Replication	Block Size	Name
0	0 B	under13
0	0 B	under18
0	0 B	under21

```
ubh01@ubh01:~$ hadoop fs -put under13.csv /user/hive/warehouse/appleproject.db/under13
22/05/22 15:06:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubh01@ubh01:~$ hadoop fs -put under21.csv /user/hive/warehouse/appleproject.db/under21
22/05/22 15:11:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubh01@ubh01:~$ hadoop fs -put under18.csv /user/hive/warehouse/appleproject.db/under18
22/05/22 15:11:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
put: under18.csv: No such file or directory
ubh01@ubh01:~$
```

Ubuntu [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Terminal Thu 23:39 ~ ubh01@ubh01:~

```
File Edit View Search Terminal Help
Query OK, 1 row affected (0.01 sec)

mysql> mysql> INSERT INTO Sony VALUES("S020", "Lakshya (title track)", "Harry thompson", "thompsonharry@gmail.com", "Rock", 43016960, 397000, 3400, "4M415"
Query OK, 1 row affected (0.00 sec)

mysql> select * from Sony;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Sony_id | Sony_title | Customer_name | Customer_email | Sony_genre | Sony_Views | Sony_Likes | Sony_Dislikes | Sony_Duration | Sony_yop |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| S001 | Kya Baat Ay | Ajit singh | ajitsingh@gmail.com | Pop | 620878106 | 4100000 | 4450 | 3M105 | 2018 |
| S002 | Raastaan Lambyan | Ram gopal | gopalram12@gmail.com | Pop | 666775437 | 6500000 | 3286 | 3M135 | 2021 |
| S003 | Genda Phool | Priyanka tripathi | ptripathi@gmail.com | Pop | 900377821 | 6200000 | 12000 | 2M535 | 2021 |
| S004 | Naah | Arjun singh | singharjung@gmail.com | Pop | 785677341 | 5600000 | 7800 | 3M225 | 2018 |
| S005 | Dyanmite | Riddhi banerjee | mnmdm123@gmail.com | Disco | 1893938 | 100000 | 4200 | 3M375 | 2020 |
| S006 | Ima's On fire | Lewis | andrewlewis123@gmail.com | Dance pop | 35000 | 181 | 14 | 4M455 | 2022 |
| S007 | Maahi ve | David peter | peterdavid@gmail.com | Rhythm | 125600274 | 681000 | 2120 | 6M115 | 2014 |
| S008 | Dance masti | Sunil chouksey | choukseysunil@gmail.com | Jazz | 20116138 | 109000 | 9887 | 5M315 | 2015 |
| S009 | Klsmat Se | Manjari shrt | shrtnanjari@gmail.com | Pop | 676815 | 5000 | 128 | 4M525 | 2014 |
| S010 | Maa da ladla | Smith johnson | johnsonsmith@gmail.com | Synthpop | 878027653 | 140000 | 5643 | 2M465 | 2014 |
| S011 | Anak tunak tun | Vittu | vittuswami123@gmail.com | Bhangra music | 1938000 | 2600000 | 10000 | 4M145 | 2014 |
| S012 | Maahi | Shubhl sharmo | shubhsharmo123@gmail.com | Blues & blues | 210150 | 35 | 94 | 4M055 | 2012 |
| S013 | Tim aaye | Ranya pandit | panditranya@gmail.com | Rhythm | 189883 | 1300 | 142 | 4M335 | 2014 |
| S014 | Barfi! (title track) | Martin lopez | lopezmartin@gmail.com | Pop | 3166931 | 14477 | 780 | 4M345 | 2012 |
| S015 | Bhaag milkha bhaag | Jack wilson | wilsonjack@gmail.com | Rock | 17288866 | 176000 | 1780 | 2M385 | 2013 |
| S016 | Raanjhana | Gaurav rajput | rajputaurav@gmail.com | Classical | 20203235 | 240000 | 2700 | 4M085 | 2013 |
| S017 | Tum tera | Martin willor | taylormartin@gmail.com | Pop | 2074800 | 202000 | 1800 | 5M035 | 2013 |
| S018 | Lata maha | Arpit peter | peterarpit@gmail.com | Pop | 252000525 | 300000 | 2200 | 3M135 | 2021 |
| S019 | Kabhi khushi kabhi gham | Raj awasthi | awasthiraj@gmail.com | Popular music | 58506600 | 353000 | 1346 | 3M135 | 2014 |
| S020 | Lakshya (title track) | Harry thompson | thompsonharry@gmail.com | Rock | 43016960 | 397000 | 3400 | 4M415 | 2015 |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
20 rows in set (0.00 sec)
```

mysql> INSERT INTO Sony VALUES("S019", "Kabhi khushi kabhi gham", "Raj awasthi", "awasthiraj@gmail.com", "Popular music", 58506600, 353000, 1346, "5M135" 2014);

INSERT INTO Sony VALUES("S020", "Lakshya (title track)", "Harry thompson", "thompsonharry@gmail.com", "Rock", 43016960, 397000, 3400, "4M415" 2015);

CREATE TABLE T-Series (TS_id VARCHAR(5), TS_title VARCHAR(50), TS_Customer_name CHAR(20), TS_Customer_email VARCHAR(30), TS_genre CHAR(20), TS_VIEWS INT(15), TS_Likes INT(10), TS_Dislikes INT(10), TS_Duration VARCHAR(10), TS_yop INT(5));

Ubuntu [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Firefox Web Browser Fri 00:10 ~

Browsing HDFS Sqoop - Google Drive Project - Google Drive +

localhost:50070/explorer.html#/user/ubh01/sqoop/apple_all_tables

Hadoop Overview Datanodes Snapshot Startup Progress Utilities -

Browse Directory

/user/ubh01/sqoop/apple_all_tables Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	ubh01	supergroup	0 B	27/5/2022, 12:07:18 am	0	0 B	Audio
drwxr-xr-x	ubh01	supergroup	0 B	27/5/2022, 12:07:35 am	0	0 B	Sony
drwxr-xr-x	ubh01	supergroup	0 B	27/5/2022, 12:07:51 am	0	0 B	T_Series
drwxr-xr-x	ubh01	supergroup	0 B	27/5/2022, 12:08:07 am	0	0 B	Video_table

Hadoop, 2015.

6.3 HADOOP CODE

```
hadoop fs -put under13.csv /user/hive/warehouse/appleproject.db/under13
hadoop fs -put under18.csv /user/hive/warehouse/appleproject.db/under18
hadoop fs -put under21.csv /user/hive/warehouse/appleproject.db/under21
hadoop fs -put Us_under_013.csv
/usr/hive/warehouse/appleproject.db/us_under_013
hadoop fs -put Us_under18.csv
/usr/hive/warehouse/appleproject.db/us_under_18
hadoop fs -put sony_data.csv /user/hive/warehouse/appleproject.db/sony_data
hadoop fs -put newmaster.csv /user/hive/warehouse/appleproject.db/newmaster
```

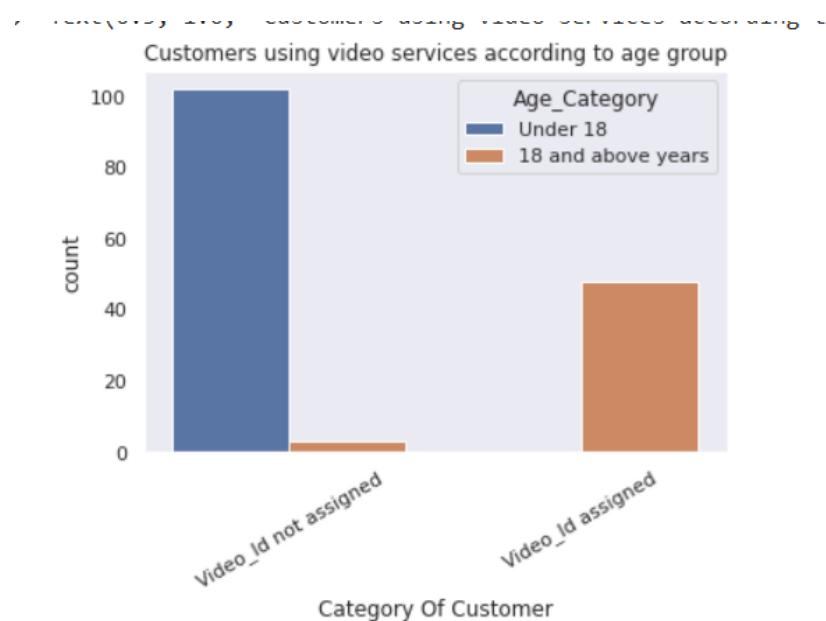
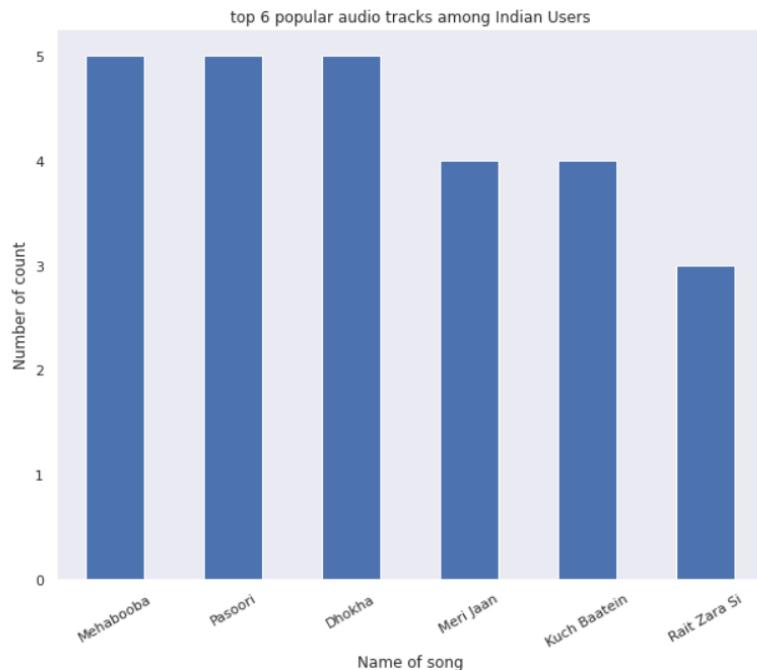
UNDER 21 QUERY

```
hive (appleproject)> create table under21( customer_id string, first_name string, last_name string, middle_name string,
> age string, audio_id string, audio_name string, product_id string, product_name string,
> music_app string, app_id string, video_id string, music_label string)
> row format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 0.185 seconds
hive (appleproject)> select * from under18 limit 10;
OK
Time taken: 0.255 seconds
```

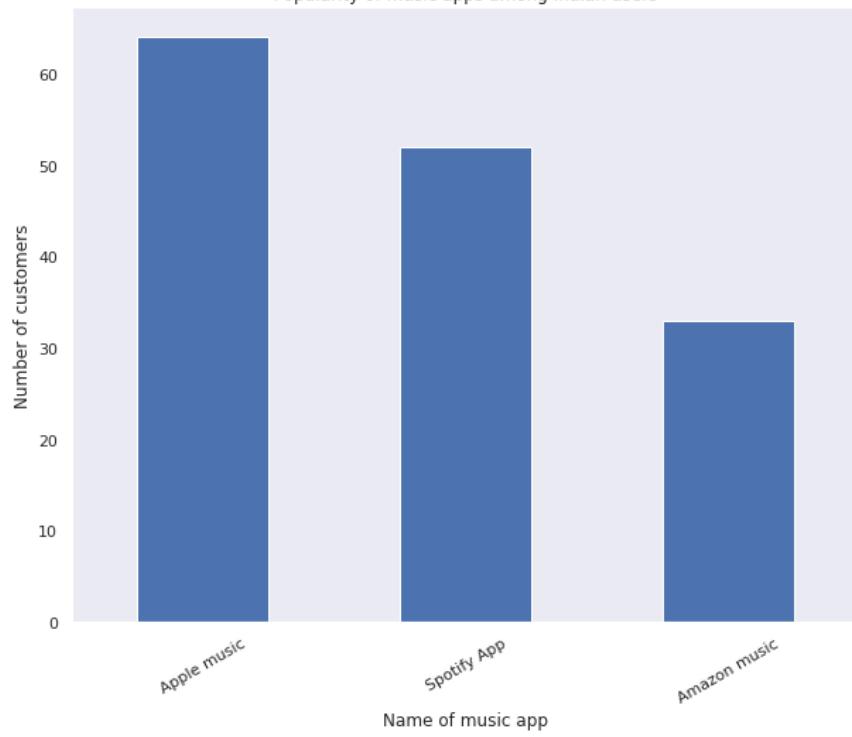
UNDER 18 QUERY

```
hive (appleproject)> select * from under18 limit 10;
OK
Customer_Id First_Name Last_Name Middle_Name Age Audio_IdAudio_Name Product_Id Product_Name Music_App App_Id Video_IdMusic_Label
CU1801 Aakash Praveen M 17 A001 Heer Raanjhana P1 Ipad Spotify MA2 Sony Music
CU1802 Amal George 16 A002 Dhokha P2 Macbook Spotify MA2 T-Series
CU1803 Aryadeep Chaudhary Praveen 17 A009 Mehbooba P2 Macbook Apple music MA3 T-Series
CU1804 Chryvash Mesare 16 A013 Meri Zindagi Hai Tu P2 Macbook Spotify MA2 T-Series
CU1805 Srijest Dhokha 17 A007 Pasoori P2 Macbook Spotify MA2 Coke Studio
CU1806 Santosh Makhtija Ameya 17 A004 Kuch Baatein P3 Iphone Apple music MA3 T-Series
CU1807 Nilesh Tiwari Suresh 16 A004 Kuch Baatein P1 Ipad Apple music MA3 T-Series
CU1808 Rajendra Mittal Ramesh 17 A030 Perfect P1 Ipad Apple music MA3 T-Series
CU1809 Prakash Munger Ram 16 A005 Jo Mujhe Deewana Kar DeP2 Macbook Amazon music MA1 T-Series
Time taken: 0.266 seconds, Fetched: 10 row(s)
hive (appleproject)>
```

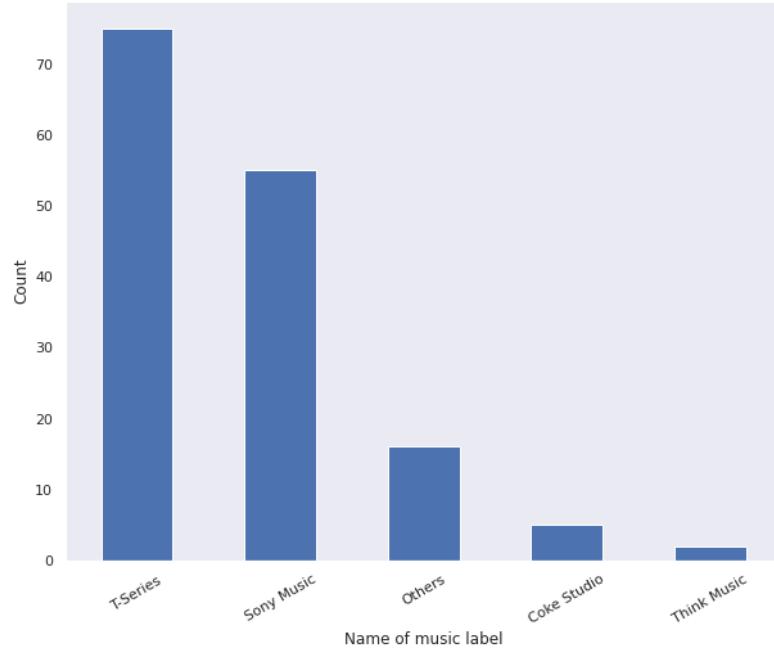
7. Output Screens



Popularity of music apps among Indian users



Top music labels preferred among Indian Customers



8. Business Benefits/Conclusion

Based on the above analysis, Apple will create a new business strategy to acquire more customers, and prevent users from switching to external party apps. The company can now understand the usage of its products based on age groups.

9. Further Enhancements/Recommendations

This project has a very vast scope in the future and even present for all app and product based companies. We developed this project for Apple but can be generalised for other contenders like Spotify, Amazon, YouTube Music, etc. There are various use cases that can be achieved by this project. Some of the future scopes are mentioned below:

- Music and video labels can use this data to learn about the genres preferred by different age groups and cater to them accordingly.
- The companies can develop algorithms in such a way that only preferred content reaches the correct audience which can result in popularity of both audio and videos.

10. References/Bibliography

- [DRAW.IO](#)
- Big Data Ethics [2]
- [GeekforGeeks](#)
- [Spark for Data Science with Python](#)
- [Spark and Python for Big Data with PySpark](#)
- [The Ultimate Hands-On Hadoop: Tame your Big Data!](#)

