Data Visualization

J Component Review 3

An Analysis of the Recent Music Trends on Online Streaming Platforms such as Spotify

Submitted by

Laya Rathod – 18BCE2162

Shreya Bage – 18BCE2291

Faculty – Priya G

Slot – F2

Course Code – CSE3020

ACKNOWLEDGEMENTS

Firstly, we would like to express our gratitude towards Mrs. Priya G. She has been supportive guide and a significant motivator behind this project. Without her guidance, this project would not have come into existence.

Secondly, we would like to thank VIT UNIVERSITY for presenting this opportunity to us. We have gained considerable knowledge and it has been a remarkable experience working together to apply our skills in a practical environment.

We would also like to thank all our classmates and fellow students in this college, for participating in our pre-implementation survey and providing us with inspiration to carry out this project.

Abstract

Streaming services have become a key player in the cultural industries in sharing media content with audiences. This chapter addresses how on-demand music-streaming services, the world's most popular format for the distribution of recorded music, have driven new professional music industry practices that are affected by, and affect in turn, the ways in which music communicates. It identifies new dynamics in the relationship between listeners and music. It then analyzes the ways in which these dynamics afford yet other distribution practices in the music industry, according to two communication patterns. These patterns have particular purposes and methods but share an alignment with the logics of distributed communication, either within or outside of the streaming services, where the struggle for audience attention is paramount. The chapter concludes with a discussion of streaming's impact upon the negotiation of new practices in the music industry derived from the abundance and intangibility of those services, as well as their multiple options for music consumption. The chapter explains how the communication adapted to the streaming paradigm is characterized by content circulation among the layers and fragments of global networks and multiple platforms, linking artists, fans, music, and the industry in new, less predictable ways. The work of communication management hence grows in importance in a streaming-dominated music industry that might also be characterized as a communication industry in its own right.

Music rankings are mainly aimed at marketing purposes but also help users in discovering new music as well as comparing songs, artists, albums, etc. This work will present an interactive way to visualize, find and compare music rankings using different techniques, as well as displaying music attributes. Our visualization makes easier to obtain information about artists and tracks, and also to compare the data gathered from the top major music rankings of Spotify.

Different visualization tools are used to show various aspects of the dataset found. This study is necessary since it helps in the recommendation algorithms that are used in modern day streaming apps for music. Song specific data is also visualized which shows the recent interests of people in music which helps artists to keep up and blend in to produce new music. It also shows which artists have been the most influential and who has given most number of hits in the past decade. The visualizations show the top genres and also the shift of the music taste over the decade.

The goal of this project is to visualize and analyse the recent trends in music over streaming applications such as Spotify for consumers to have a better idea of the popularity of artists as well as the music taste depending on the timelines which can be very useful in recommendation algorithms used in future streaming platforms.

Introduction

The term ‘streaming media’ is ambiguous and, generally speaking, used rather loosely in everyday and industrial contexts. The forces that brought about streaming have been recounted by others (Hesmondhalgh and Meier, 2017) and are beyond the scope of detail than can be recounted here. Historically, ‘streaming’ appears in the 1990s to describe a technical process for delivering media over the internet in ‘real time,’ without the file being downloaded or stored on a local drive. Alternatively, the phrase sometimes refers to forms of ‘on-demand’ services regardless of the technical means of transmission, such as cable video on-demand, and it is possible that viewers might conceive of catch up services (US MVPD video on demand) as streaming as well. More and more, however, ‘streaming’ refers to a particular kind of media service that is increasingly mainstream in music, movies and television. Key features of these services include the availability of subscription payment for on-demand access to a large media catalogue over internet protocols - though we cannot claim to be focusing on ‘subscription streaming services’ given that most music streaming occurs through free versions of these services. The streaming video ecosystem, too, is much broader than the focus here, and the industrial dynamics among various sectors make common claims difficult. In order to make this exploratory paper manageable, we will primarily concentrate on the two largest services in movie, television and music spheres: Netflix and Spotify.

The history of music technology is peppered with breakthroughs that freed people from constraints. Radio meant no longer physically having to go to a live performance if you wanted to listen to music. Vinyl gave people the option of playing records they owned rather than being stuck listening to whatever was on the radio at the time. Then cassettes made a personal music collection portable, so you could listen to it anywhere.

We’re now deep in the digital era, untangling recorded music’s historical link with physical media – and things are changing at speed. Just 15 years ago, people were excited to own a device the size of a deck of cards that stored a thousand songs. Now, countless pieces of internet-connected hardware offer instant access to tens of millions of music tracks.

Modern streaming services – TIDAL, Spotify, Apple Music *et al* – are the culmination of two decades of cutting-edge technological achievements, evolution and iteration. Their architects have unpacked and developed audio file formats, delivery platforms and smart algorithms, and sharp-eyed businesspeople have converted that vision into the subscription models we know today.

Spotify is a digital music, podcast, and video streaming service that gives you access to millions of songs and other content from artists all over the world. It offers streaming of over 30 million songs. The service was founded in 2006 by Daniel Ek and Martin Lorentzon. Spotify has around 140 million active registered users. There are two versions of Spotify: a premium monthly subscription service and a free service which is supported by advertising. It is available in more than 75 countries and is considered one of the best and most famous music streaming service in the world.

Africa	Algeria, Egypt, Morocco, South Africa, Tunisia.
Asia	Bahrain, Hong Kong, India, Indonesia, Israel, Japan, Jordan, Kuwait, Lebanon, Malaysia, Oman, Palestine, Philippines, Qatar, Saudi Arabia, Singapore, Taiwan, Thailand, United Arab Emirates, Vietnam.
Europe	Andorra, Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Spain, Sweden, Switzerland, Turkey, United Kingdom.
North America	Canada, Costa Rica, Dominican Republic, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, United States.
South America	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay.
Oceania	Australia, New Zealand.

In today's millennial world, most people listen to their music on streaming platforms. There is a huge demand and fanbase for music. There are so many different genres and music to choose from and everyone has a different taste for music choices. This project has a great scope as it visualizes all the recent trends in music over the past decade based on the data collected on Spotify which is the leading music streaming service in the world today. It considers all the top ranking songs and artists based on year for it visualization. This analysis is very relevant and needed since all Machine learning based recommendation algorithms are based on the past trends that have followed in music releases.

Literature Survey/ Related work

Analyzing the Spotify Top 200 Through a Point Process Lens Michelangelo Harris, Brian Liu, Cean Park, Ravi Ramireddy, Gloria Ren, Max Ren, Shangdi Yu, Andrew Daw, and Jamol Pender∗ School of Operations Research & Information Engineering Cornell University, Ithaca, NY, October 4, 2019

Learning about Spotify the world's largest music streaming service, streaming audio is a great way of delivering sound without having to download files from the internet. This paper takes a closer and more specific venture into Spotify's popularity. It begins ab expedition through an analysis of the Spotify Top 200, a chart of the 200 most frequently streamed songs each day on Spotify. This investigation is built on a data set scraped from over 20 months of the United States based edition of these rankings, spanning from 2017 to 2018. As defined by Spotify, one day spans from 3:00 PM UTC through 2:59 PM UTC on the next. We have scraped this data set from the publicly available Spotify Top 200 charts, specifically from the U.S. based rankings. The code to do so is derived from open source code used to form a Kaggle data set containing the worldwide Top 200 rankings for all of 2017 [2]. We also learnt from this paper hoe quantify the duration of a pop hit.

Approaching Media Industries Comparatively: A Case Study of Streaming, Daniel Herbert, Amanda Lotz and Lee Marshall

This paper got us to understand the concept of streaming data. We understood what is the importance for streaming data in different multimedia platforms. Key features of these services include the availability of subscription payment for on-demand access to a large media catalogue over internet protocols - though we cannot claim to be focusing on 'subscription streaming services' given that most music streaming occurs through free versions of these services. This paper gave us the analysis of business practices related to streaming of audio. We learnt that suggesting that implications for each industry derive from pre-streaming norms at least as much as the common implications of streaming. There is notable consistency in the consumer experience across streaming media. The experience of these three industries in conversation revealed that the timeline of adjustment of different components such as consumer experience, business practices and content is occurring at different rates in each medium.

Spotify Teardown: Inside the Black Box of Streaming Music by Maria Eriksson, Rasmus Fleischer, Anna Johansson, Pelle Snickars & Patrick Vonderau, The MIT Press, ISBN 978-0262038904 (Paperback)

In try-ing to explore the way in which Spotify is commonly perceived, the re-searchers analyse Spotify from every possible angle. Chapter 1 "Where Is Spotify" starts by questioning Spotify's corporate history, analysing it through the lens of its funding rounds. This approach proves helpful to understand the way the company was built from a service that relied on pirated material. The second chapter "When Do Files Become Music" explores the system's backstage describing its network infrastructure and the data gathering and sharing processes. One of the chapters also talk about how recommendations are fed to users based on changing circumstances. The final chapter talked about the financial tactics of the company.

Infinite content and interrupted listening The impact of smartphones, streaming and music ‘superabundance’ on everyday personal music listening behavior, Ellen Moore

Music is something that has been around for almost as long as we can remember. This paper however has a very unique way of understanding and interpretation done by the author. The intention of this research was to produce new insights into everyday listening and our relationship to music in the age of smartphones and the infinite content provided by streaming services. A study was conducted to see if the youth today can survive without the wonderful integration of music and technology. Many of the teens couldn’t. And that automatically brings us to the importance of streaming audio.

Metrics and decision making in music streaming, Arnt Maaso & Anju Nyland Hagen

As we have seen, over the last few years, streaming services, and Spotify in particular, have made an increasing amount of data available to different stakeholders on a daily basis, and those stakeholders have enjoyed increasing flexibility to monitor and act based upon these Metrics. No single stakeholder has access to all relevant data points about music, but it is clear that the MSSs and major labels are dominating the metrics race. These actors also have the resources and skillsets to interpret this data in ways that others do not. When metrics are presented, interpreted and analyzed upon by numerous stakeholders making decisions that is the fed back to algorithms that create consumer software.

Music in Streams: Communicating Music in the Streaming Paradigm Anja Nylund Hagen, Postdoctoral fellow, Dept. of Musicology, University of Oslo

This paper got us into the analysis of how streaming shapes music communication. Music can now be run on the smartphones on the go as well as randomly. How songs are marketed on these sites are also discussed artists are like brands for these companies.

Music Radio as a Format Remediated for Stream-based Music, Andreas Lenander Egidius

This paper drove into the concept of radio. Talked about how spotify and apple music technologies have taken it to a whole different level. Both providing users with endless new features signifying a huge evolution from the radio.

Musical trends and predictability of success in contemporary songs in and out of the top charts, (Royal society of Science)

This is paper is similar to what we will be doing for our project. We will analyze the given dataset that is the songs releases in a particular country. The influence of the musical characteristics were observed in how it managed to change the trends on the charts.

Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery, Hannes Datta, George Knox, Bart J. Bronnenberg*

Talks about a different perspective of live streaming. In recent years, copyright-related industries have suffered as new digital technologies disrupted their revenue models. One such disruptive technology that is currently taking over the music industry is streaming. Streaming allows consumers unlimited access to a vast library of content at a fixed monthly payment.

The rise of streaming music and implications for music production, R. Scott Hiller∗ Jason Walter†‡, July 11, 2016

In this paper we observed that depending on the mounting popularity of the non-durable option of streaming. With a sufficiently low price and sufficiently high base of listening, the paid streaming model becomes an attractive option for consumers. No longer forced to pay for individual durable bundles, a subscription allows access to a much larger, diverse bundle.

<u>Proposed Methodology/ Project Description</u>

Firstly, a literature survey is being compiled where we have read several research papers on streaming platforms, Spotify as an application, recent music trends, recommendation algorithms, etc. We have built a problem statement based on the content that we found in these articles. Spotify has its own website where many research articles and blogs have been posted.

Secondly, we identify the most appropriate dataset for our requirements from dataset repository sites such as <u>www.kaggle.com/datasets</u>.

Thirdly, we use visualization tools such as Tableau to visualize various aspects of the dataset in the most suitable forms. Spotify Labs and various other online articles have given demo visualization techniques to portray such data. We will do a research on what people want to know about this field and try to visualize that.

Next, we analyze the data and devise conclusions from them about the trends of the past decade. In the end we summarize all the results that we found.

What do people want to know about the music trends over the last decade?
The artists, songs as well as the specifications of each song.

Why do people want to know about it?
To see which artist performed well and which songs were ranked highest in each year and also the type and genre of the songs which changed over the decade depending upon the music preferences of the audience.

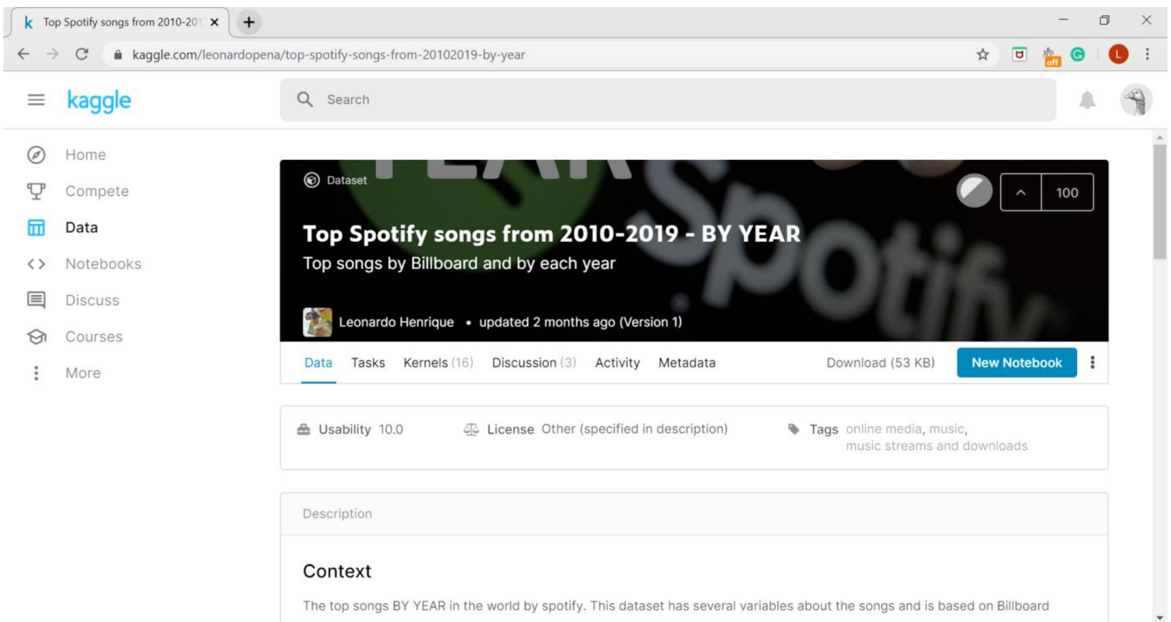
How will we visualize the data?
Using datasets available from Spotify and Kaggle.com and by using tools such as Tableau and Scikit-Learn in Python to visualize the trends in various forms like scatter plots, area plots, bar charts, pie charts, line charts etc.

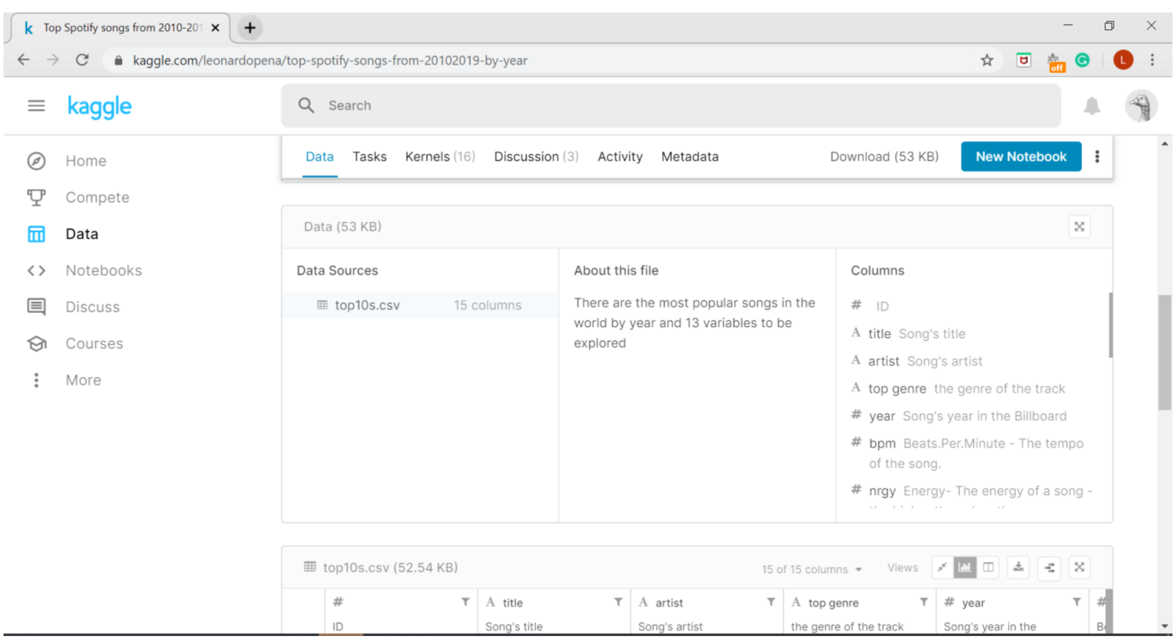
At Last, we use Machine Learning Algorithms such as K-Means Clustering using Scikit Learning in Python to make playlists of similar songs using our analysis. The main use of these analysis is in recommendation systems in music streaming apps.

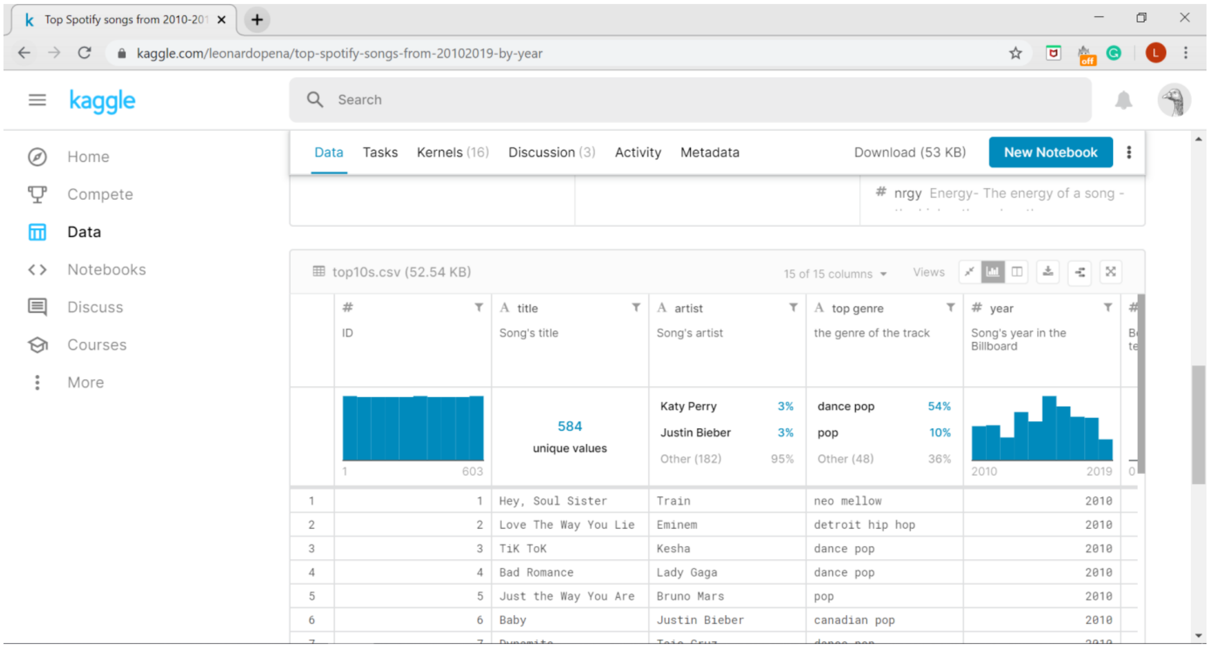
Dataset

Our first task was to obtain a large set of songs. After a bit of research, we decided that the **Top Spotify songs from 2010-2019 – By Year (Top songs by the Billboard and by each year)** dataset would best suit our needs for this project since we could download an excel file of track names and artists which we could then use in Tableau. Our main challenge here was to select the right amount and range of songs. Data could become heavily skewed at this initial step. We wanted to sample as many genres as possible in order to create an accurate high-level view of musical changes over time. We also needed to ensure our data had a sufficient number of songs from all time periods for accurate analysis.

Getting the Dataset from kaggle.com/datasets?search=spotify







There are a total of 603 songs in the dataset.
It has the following columns which represent the different attributes of each song.

1. **# ID**
2. **title :**Song's title [Categorical]
3. **artist :**Song's artist [Categorical]
4. **top genre :**the genre of the track [Categorical]
5. **# year :**Song's year in the Billboard [Quatitaive]
6. **# bpm :**Beats.Per.Minute - The tempo of the song. [Quatitaive]
7. **# nrgy :**Energy- The energy of a song - the higher the value, the more energtic. Song [Quatitaive]
8. **# dnce :**Danceability - The higher the value, the easier it is to dance to this song. [Quatitaive]
9. **# dB :**Loudness..dB.. - The higher the value, the louder the song [Quatitaive]
10. **# live :**Liveness - The higher the value, the more likely the song is a live recording [Quatitaive]
11. **# val :**Valence - The higher the value, the more positive mood for the song. [Quatitaive]
12. **# dur :**Length - The duration of the song. [Quatitaive]
13. **# acous :**Acousticness.. - The higher the value the more acoustic the song is. [Quatitaive]
14. **# spch :**Speechiness - The higher the value the more spoken word the song contains. [Quatitaive]
15. **# pop :**Popularity- The higher the value the more popular the song is. [Quatitaive]

Tools used for Visualizations

Tableau

Tableau is a powerful data visualization and discovery tool. It is an important part of a data analyst or data scientist's – skill set, with many organizations specifying it as a key skill in job adverts.

In this article, we'll take a look at few things in Tableau you need to know to successfully make a mark in your business intelligence career.

While architecture of traditional BI tools has hardware limitations, Tableau does not have such dependencies and it can function independently and requires minimum hardware support. Traditional tools are based on a complex set of technologies when Tableau is based on Associative Search technology making it intuitive, fast and dynamic. Tableau supports in-memory, multi-thread and multi-core computing and more advanced capabilities while traditional BI tools do not offer such functionalities.

Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It helps in simplifying raw data into the very easily understandable format.

Data analysis is very fast with Tableau and the visualizations created are in the form of dashboards and worksheets. The data that is created using Tableau can be understood by professional at any level in an organization. It even allows a non-technical user to create a customized dashboard.

The best feature Tableau are:

- Data Blending
- Real time analysis
- Collaboration of data

The great thing about Tableau software is that it doesn't require any technical or any kind of programming skills to operate. The tool has garnered interest among the people from all sectors such as business, researchers, different industries, etc.

- What is Tableau?
- Tableau Desktop
- Tableau Public
- Tableau Server
- Tableau Online
- Tableau Reader
- How does Tableau work?
- Excel Vs. Tableau

Tableau Product Suite

The Tableau Product Suite consists of:

- Tableau Desktop
- Tableau Public

- Tableau Online
- Tableau Server
- Tableau Reader

How does Tableau work?

Tableau connects and extracts the data stored in various places. It can pull data from any platform imaginable. A simple database such as an excel, pdf, to a complex database like Oracle, a database in the cloud such as Amazon webs services, Microsoft Azure SQL database, Google Cloud SQL and various other data sources can be extracted by Tableau.

When Tableau is launched, ready data connectors are available which allows you to connect to any database. Depending on the version of Tableau that you have purchased the number of data connectors supported by Tableau will vary.

The pulled data can be either connected live or extracted to the Tableau's data engine, Tableau Desktop. This is where the Data analyst, data engineer work with the data that was pulled up and develop visualizations. The created dashboards are shared with the users as a static file. The users who receive the dashboards views the file using Tableau Reader.

The data from the Tableau Desktop can be published to the Tableau server. This is an enterprise platform where collaboration, distribution, governance, security model, automation features are supported. With the Tableau server, the end users have a better experience in accessing the files from all locations be it a desktop, mobile or email.

- Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry
- The Tableau Product Suite consists of 1) Tableau Desktop 2) Tableau Public 3) Tableau Online 4) Tableau Server and Tableau Reader
- Tableau Desktop has a rich feature set and allows you to code and customize reports
- In Tableau public, workbooks created cannot be saved locally, in turn, it should be saved to the Tableau's public cloud which can be viewed and accessed by anyone
- Tableau server is specifically used to share the workbooks, visualizations that are created in the Tableau Desktop application across the organization
- Tableau online has all the similar functionalities of the Tableau Server, but the data is stored on servers hosted in the cloud which are maintained by the Tableau group.
- Tableau Reader is a free tool which allows you to view the workbooks and visualizations created using Tableau Desktop or Tableau Public.
- Tableau connects and extracts the data stored in various places. It can pull data from any platform imaginable.
- The spreadsheet application used for manipulating the data while Tableau is a perfect visualization tool used for analysis

Scikit-learn in Python

Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Scikit-learn is largely written in Python, and uses numpy extensively for high-performance linear algebra and array operations. Furthermore, some core algorithms are written in Cython to improve performance. Support vector machines are implemented by a Cython wrapper around LIBSVM; logistic regression and linear support vector machines by a similar wrapper around LIBLINEAR. In such cases, extending these methods with Python may not be possible.

Scikit-learn integrates well with many other Python libraries, such as matplotlib and plotly for plotting, numpy for array vectorization, pandas dataframes, scipy, and many more.

K-Means Clustering Algorithm

The K-Means algorithm clusters data by trying to separate samples in n groups of equal variance, minimizing a criterion known as the inertia or within-cluster sum-of-squares . This algorithm requires the number of clusters to be specified. It scales well to large number of samples and has been used across a large range of application areas in many different fields.

The k-means algorithm divides a set of N samples X into K disjoint clusters , each described by the mean of the samples in the cluster. The means are commonly called the cluster "centroids"; note that they are not, in general, points from , although they live in the same space.

The K-means algorithm aims to choose centroids that minimise the inertia, or within-cluster sum-of-squares criterion:

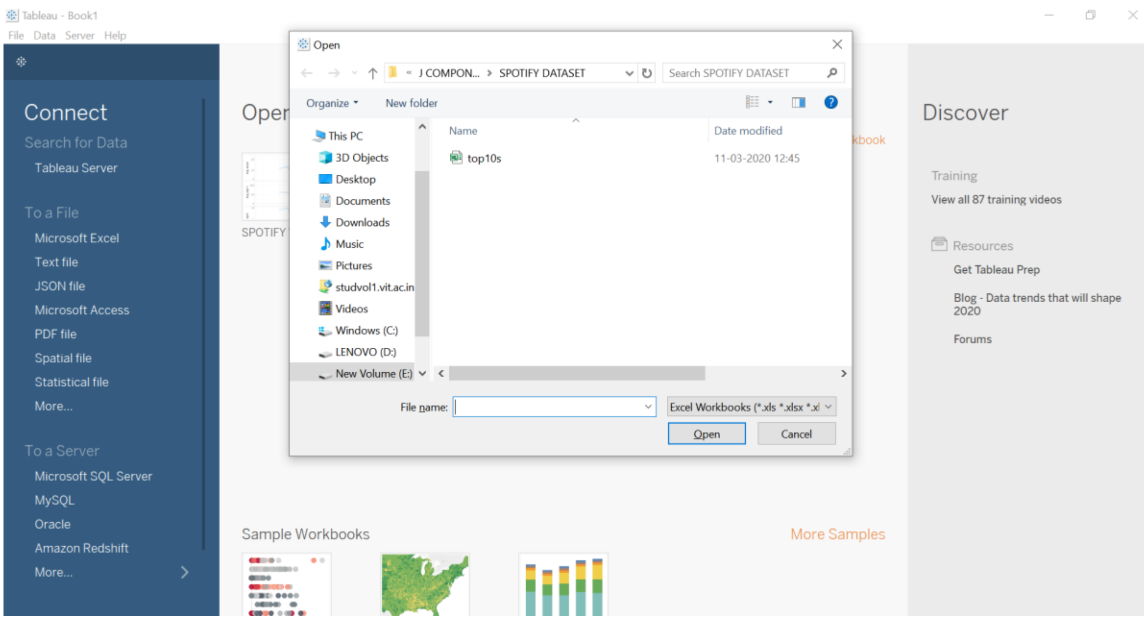
$$\sum_{i=0}^{n} \min_{\mu_j \in C}(||x_i - \mu_j||^2)$$

K-means is often referred to as Lloyd's algorithm. In basic terms, the algorithm has three steps. The first step chooses the initial centroids, with the most basic method being to choose samples from the dataset . After initialization, K-means consists of looping between the two other steps. The first step assigns each sample to its nearest centroid. The second step creates new centroids by taking the mean value of all of the samples assigned to each previous centroid. The difference between the old and the new centroids are computed and the algorithm repeats these last two steps until this value is less than a threshold. In other words, it repeats until the centroids do not move significantly.

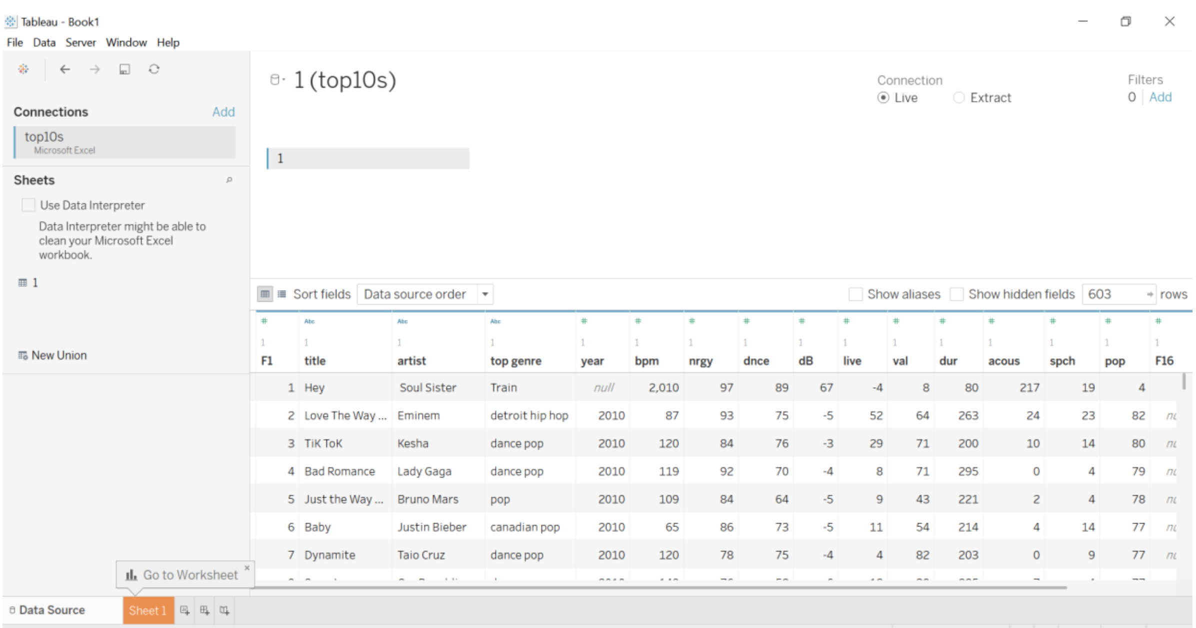
Data Pre-processing

Importing the excel file on Tableau

Click on the Microsoft excel option on the left hand side bar under Connect and To a File



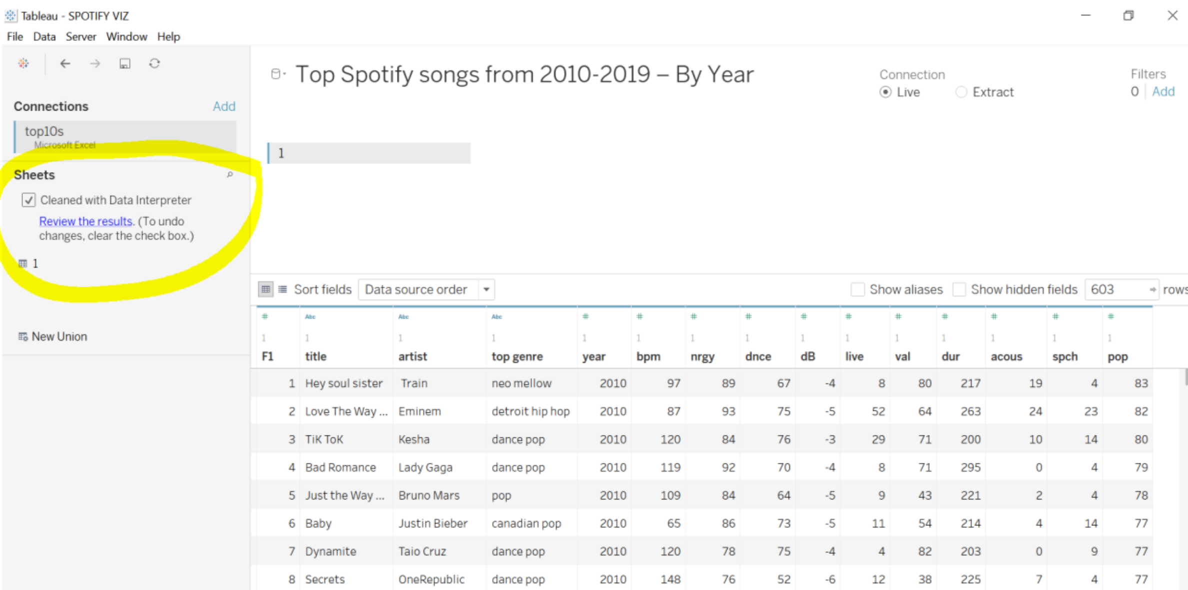
A new book opens with the file loaded. Here we can see the list of all the songs and the different attribute columns. Here we can also select which columns to be included and which to be removed.



After getting the song set, we had to clean the data so that our Spotify searches would be more successful.Firstly, we changed the name of the dataset to Top Spotify songs from 2010-2019 – By Year.

We did not include the featured artist of the songs as a separate entity eg.Things like (LP Version) and artist features (eg. feat Rihanna) were removed to make the searches more generic. They are only mentioned in the title. We have also removed the rows which contained any NA or null values for more accuracy and removed the F15 and F16 columns from the dataset which were not required for our study.

We used the Cleaned with Data Interpreter option of Tableau for the pre-processing of our dataset. This can be done by selecting the checkbox on the left-hand side panel.



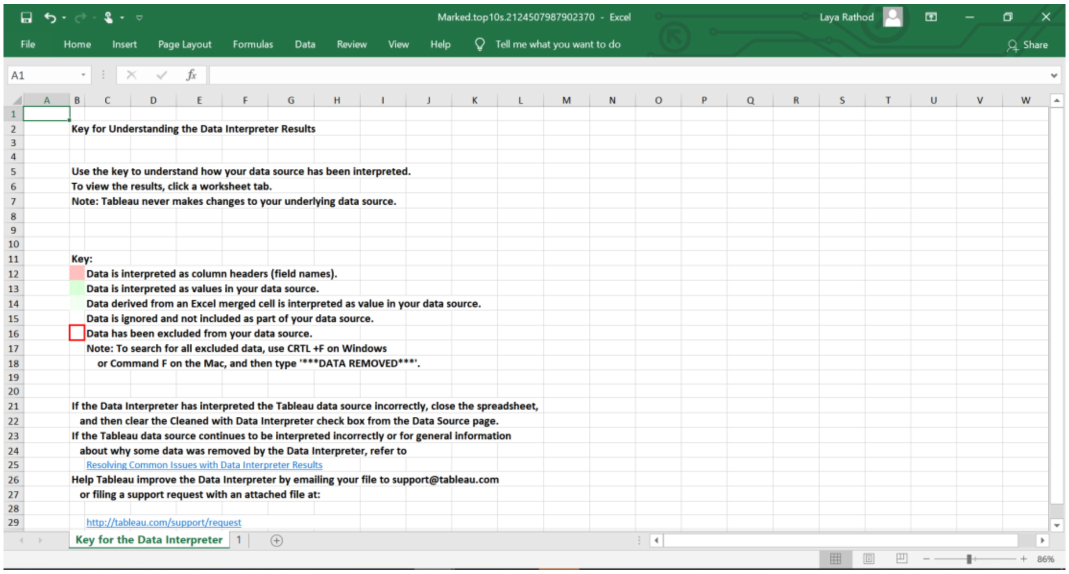
<u>What does Data Interpreter do?</u>

Data Interpreter can give you a head start when cleaning your data. It can detect things like titles, notes, footers, empty cells, and so on and bypass them to identify the actual fields and values in your data set.

It can even detect additional tables and sub-tables so that you can work with a subset of your data independently of the other data.

After Data Interpreter has done its magic, you can check its work to make sure it captured the data that you wanted and identified it correctly. Then, you can make any necessary adjustments.

After you select the data that you want to work with, you might also need to do some additional cleaning steps like pivoting your data, splitting fields, or adding filters to get the data in the shape you want before starting your analysis.



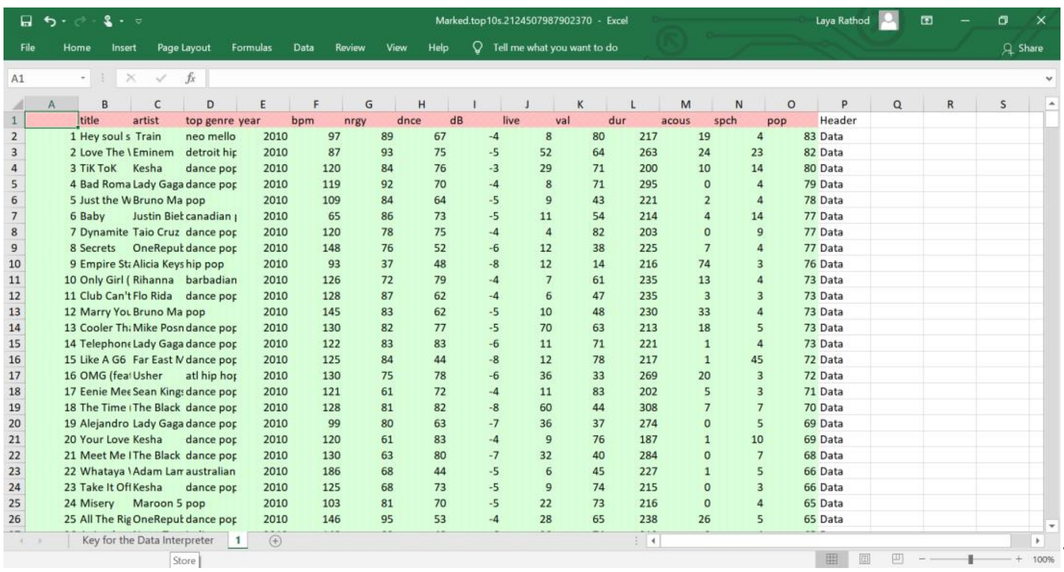
Key for Understanding the Data Interpreter Results

Use the key to understand how your data source has been interpreted.
To view the results, click a worksheet tab.
Note: Tableau never makes changes to your underlying data source.

Key:
Data is interpreted as column headers (field names).
Data is interpreted as values in your data source.
Data derived from an Excel merged cell is interpreted as value in your data source.
Data is ignored and not included as part of your data source.
Data has been excluded from your data source.
Note: To search for all excluded data, use CRTL +F on Windows
or Command F on the Mac, and then type '***DATA REMOVED***'.

If the Data Interpreter has interpreted the Tableau data source incorrectly, close the spreadsheet,
and then clear the Cleaned with Data Interpreter check box from the Data Source page.
If the Tableau data source continues to be interpreted incorrectly or for general information
about why some data was removed by the Data Interpreter, refer to
Resolving Common Issues with Data Interpreter Results
Help Tableau improve the Data Interpreter by emailing your file to support@tableau.com
or filing a support request with an attached file at:

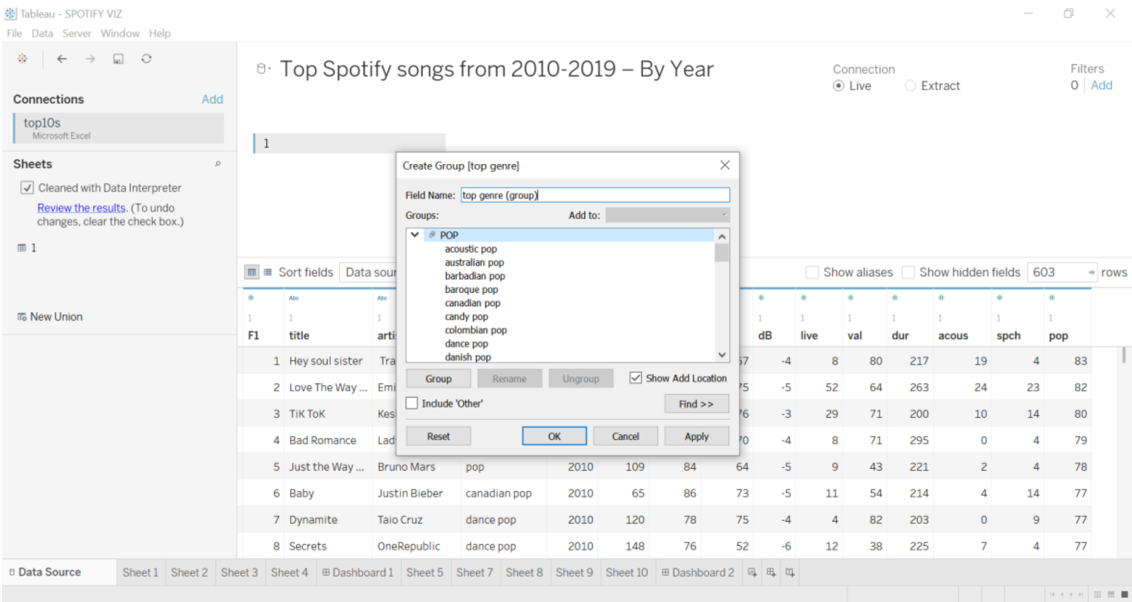
http://tableau.com/support/request

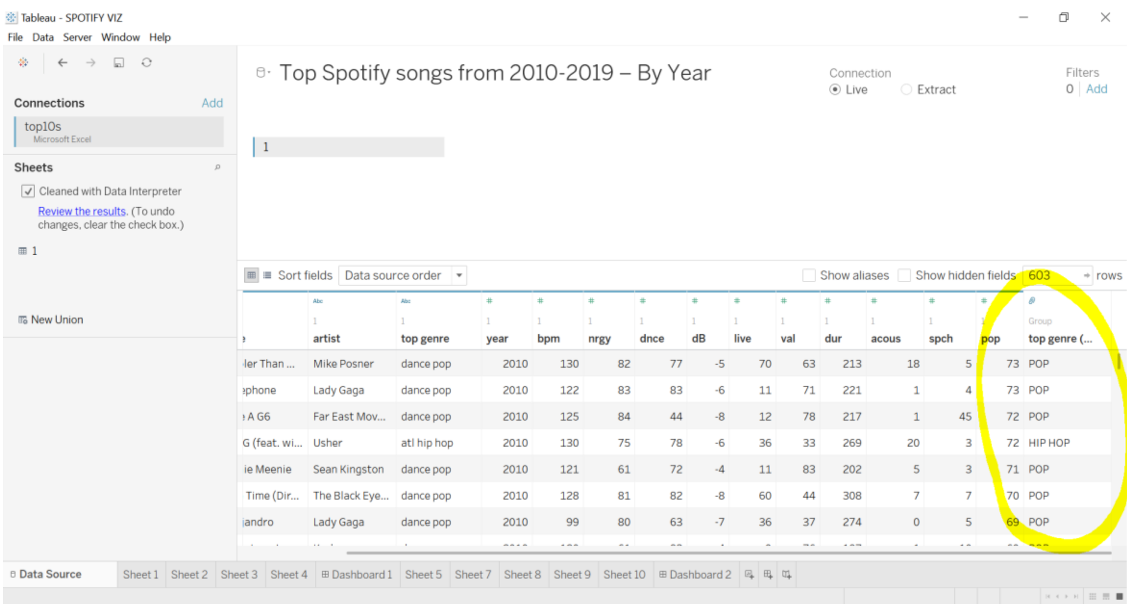


	title	artist	top genre	year	bpm	nrgy	dnce	dB	live	val	dur	acous	spch	pop	Header
1	Hey soul s	Train	neo mello	2010	97	89	67	-4	8	80	217	19	4	83	Data
2	Love The \	Eminem	detroit hip	2010	87	93	75	-5	52	64	263	24	23	82	Data
3	TiK ToK	Kesha	dance pop	2010	120	84	76	-3	29	71	200	10	14	80	Data
4	Bad Roma	Lady Gaga	dance pop	2010	119	92	70	-4	8	71	295	0	4	79	Data
5	Just the W	Bruno Ma	pop	2010	109	84	64	-5	9	43	221	2	4	78	Data
6	Baby	Justin Biet	canadian p	2010	65	86	73	-5	11	54	214	4	14	77	Data
7	Dynamite	Taio Cruz	dance pop	2010	120	78	75	-4	4	82	203	0	9	77	Data
8	Secrets	OneRepub	dance pop	2010	148	76	52	-6	12	38	225	7	4	77	Data
9	Empire Sta	Alicia Keys	hip pop	2010	93	37	48	-8	12	14	216	74	3	76	Data
10	Only Girl (	Rihanna	barbadian	2010	126	72	79	-4	7	61	235	13	4	73	Data
11	Club Can't	Flo Rida	dance pop	2010	128	87	62	-4	6	47	235	3	3	73	Data
12	Marry You	Bruno Ma	pop	2010	145	83	62	-5	10	48	230	33	4	73	Data
13	Cooler Tha	Mike Posn	dance pop	2010	130	82	77	-5	70	63	213	18	5	73	Data
14	Telephone	Lady Gaga	dance pop	2010	122	83	83	-6	11	71	221	1	4	73	Data
15	Like A G6	Far East M	dance pop	2010	125	84	44	-8	12	78	217	1	45	72	Data
16	OMG (feat	Usher	atl hip hop	2010	130	75	78	-6	36	33	269	20	3	72	Data
17	Eenie Mee	Sean Kings	dance pop	2010	121	61	72	-4	11	83	202	5	3	71	Data
18	The Time	The Black	dance pop	2010	128	81	82	-8	60	44	308	7	7	70	Data
19	Alejandro	Lady Gaga	dance pop	2010	99	80	63	-7	36	37	274	0	5	69	Data
20	Your Love	Kesha	dance pop	2010	120	61	83	-4	9	76	187	1	10	69	Data
21	Meet Me I	The Black	dance pop	2010	130	63	80	-7	32	40	284	0	7	68	Data
22	Whataya \	Adam Lam	australian	2010	186	68	44	-5	6	45	227	1	5	66	Data
23	Take It Off	Kesha	dance pop	2010	125	68	73	-5	9	74	215	0	3	66	Data
24	Misery	Maroon 5	pop	2010	103	81	70	-5	22	73	216	0	4	65	Data
25	All The Rig	OneRepub	dance pop	2010	146	95	53	-4	28	65	238	26	5	65	Data

”

We also grouped the sub genre's into more broader genre's eg: acoustic pop, Australian pop, barbadian pop were all grouped as pop and so on.





<u>For our Python implementation of K-Means clustering:</u>

- We removed all the songs which had N/A values or 0's in all rows.
- We filtered our songs according to the required years.
- We scaled/normalized the values of the required columns used in the algorithm for optimum results.

Analysis of how streaming shapes music communication

Spotify the largest on-demand music service in the world, has a history of pushing technological boundaries and using big data, artificial intelligence and machine learning to drive success. The digital music company with more than 100 million users has been busy this year enhancing its service and tech capabilities through several acquisitions. The industry watch dogs predict the company will launch an IPO in 2018.

Data: Powerful By-product of Streaming Music

When you have tens of millions of people listening to music every minute of the day, you have access to an extraordinary amount of intel that includes what songs get the most play time, to where listeners are tuning in from and even what device they are using to access the service. There's no doubt Spotify is a data-driven company and it uses the data in every part of the organization to drive decisions. As the service continues to acquire data points, it's using that information to train the algorithms and machines to listen to music and extrapolate insights that impact its business and the experience of listeners.

One example is the Discover Weekly feature on Spotify that reached 40 million people in its first year. Every user gets a personalized playlist every week from Spotify of music that they have not heard before on the service, but that will be something the listener is expected to enjoy—a modern-day version of a best friend creating a personalized mix tape.

Spotify for Artists

In an effort to make its mountains of data available to musicians and their managers, Spotify just launched the Spotify for Artists app that provides mobile access to analytics—everything from which playlists are generating new fans to how many streams they are getting overall. Think Google Analytics for musicians. It was originally launched in a web version earlier this year, but the mobile app allows musicians to access the info from the tour bus and the geographic streaming data can be instrumental to musicians and their teams to plan tours more effectively. Artists also have more control over their presence on Spotify including selecting the "artist's pick," and they can update their bios and post playlists.

Spotify Acquires Technology Firms to Enhance Service

With the acquisition of Niland, the fourth acquisition for 2017, Spotify will use the API-based product and machine learning to provide its users with better search and recommendations to help them discover music they will like.

Earlier this year, Spotify acquired the blockchain startup Mediachain Labs to help develop solutions via a decentralized database to better connect artists and licensing agreements with the tracks on Spotify's service. MightyTV, a content recommendation service, and audio detection startup Sonalytic were also acquired this year

.

What's Next for Spotify?

When news broke that Francois Pachet, a French scientist and expert on music composed by AI, joined the Spotify team to "focus on making tools to help artists in their creative process," not everyone believed that's ALL that he'd do. You can just imagine how a leader in AI (Artificial Intelligence)might use his expertise to turn the tables at Spotify to make AI-composed music that would push out artists and their labels. So far, Spotify denies that this will be the case even though this isn't the first AI feature they launched—AI Duet released earlier this year where listeners could create a duet with a computer.

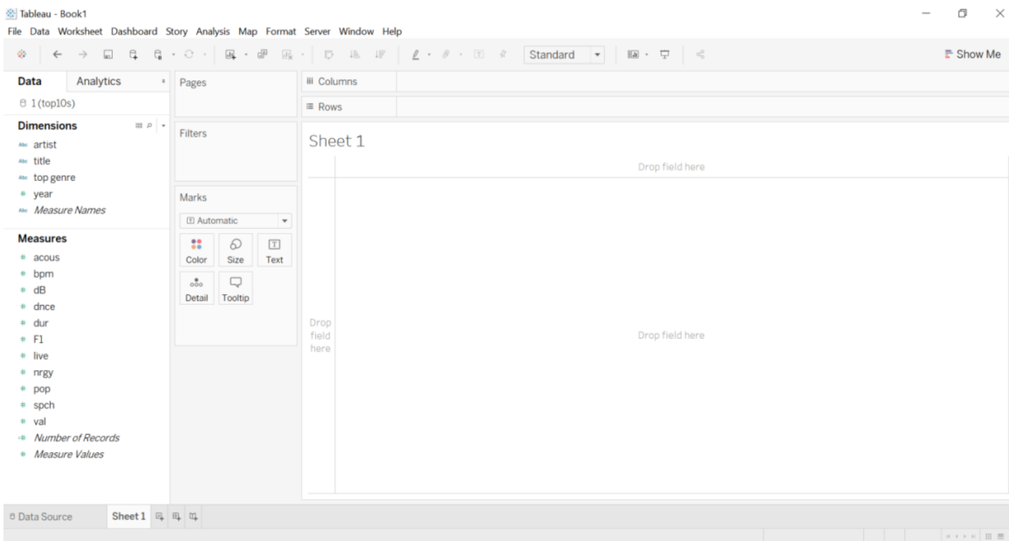
We can also expect the company to continue to humanize data in creative ways like it did when it used its vast amounts of data to launch a global ad campaign that highlighted some of the more bizarre user habits of 2016. Headlines included "Dear person who played 'Sorry' 42 times on Valentine's Day, what did you do?" and "Dear 3,749 people who streamed 'It's the End of the World as We Know It' the day of the Brexit vote, hang in there."

As Spotify learned in 2015, its community will respond if it feels like it's taking too many liberties with data. After introducing large-scale changes to its privacy policy, users let the company know they were angry by cancelling subscriptions and taking to social media to express their dismay. This prompted Spotify CEO Daniel Ek to apologize for unclear communication and made it clear any access to personal data would only occur with the permission of the individual.

We might not know today where Spotify will innovate next, but we will be watching. As innovators they will encounter learning experiences and even failures as they use big data, AI and machine learning to drive success. Those are experiences we can all learn from.

Creating Different Visualizations

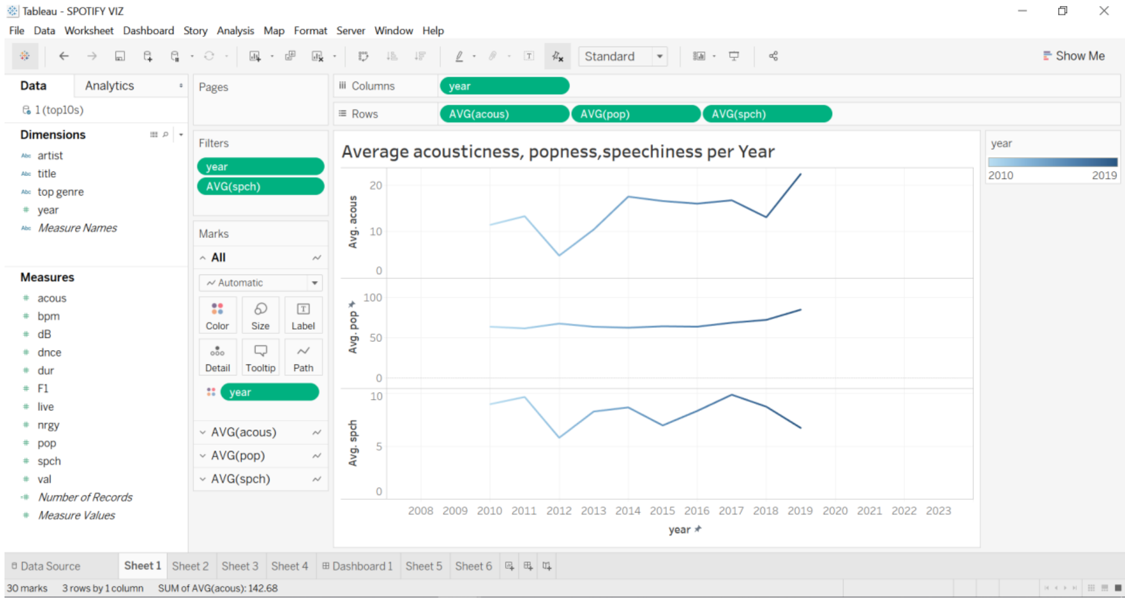
This is a blank sheet on which we can work on for a single visualization on Tableau:



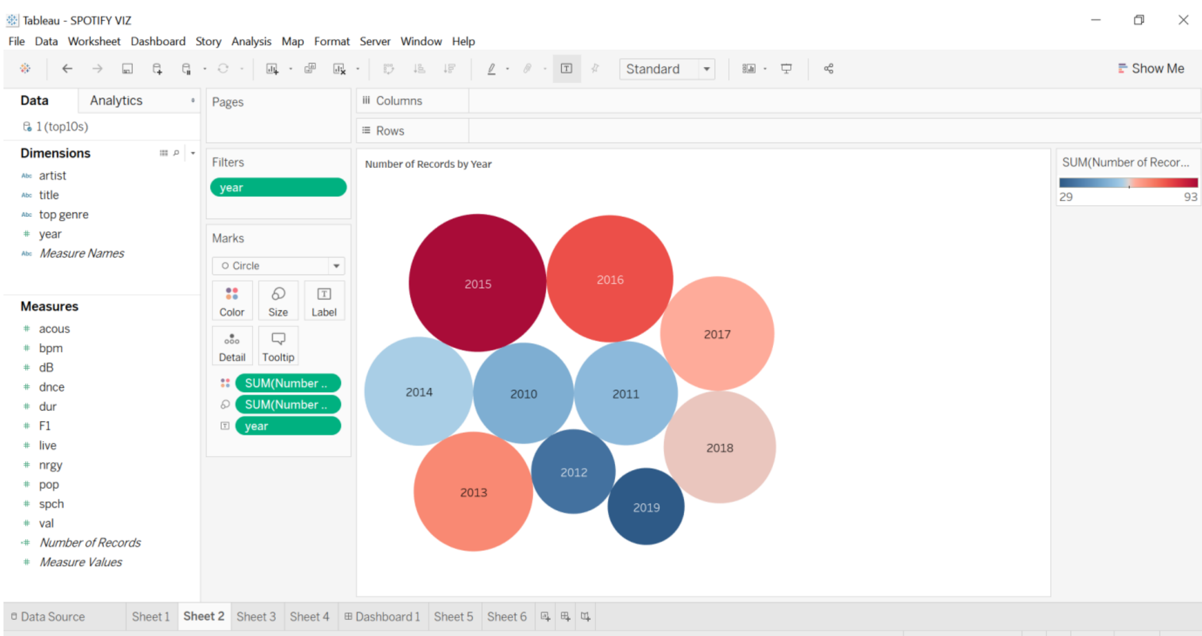
We add Dimensions and Measures from the list provided on the left- hand side to the columns and rows at the top by dragging each field. Dimensions are the Categorical attributes, while Measures are the quantitative attributes.Then we decide which type of visualization is required whose options are seen on the left hand-side panel. We can also change which kind of field we require. We can choose which kind of operation we need to visualize like sum, average, median, etc.

Then we can drag the measures and dimensions to the color marks to color map it and their we can edit the type of mapping/encoding we want. Various other identities can also be altered like size, text, detail etc. On the right- hand side we will be able to see the legends and keys for each visualization. We can add filters on the fields where we can chose which part of the visualization is visible. There are highlighters which are used to highlight specific years, artists or genre's. On hovering we can see the exact values as well as other details about the point being mapped.

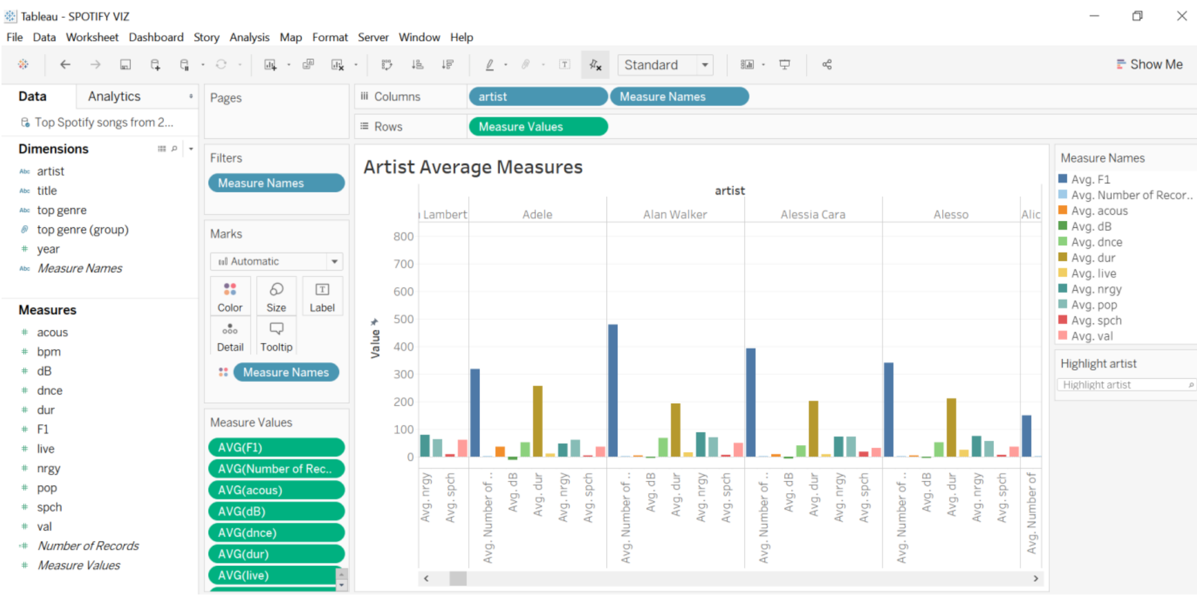
<u>Some of the visualizations created are and the Analysis that we made from it:</u>



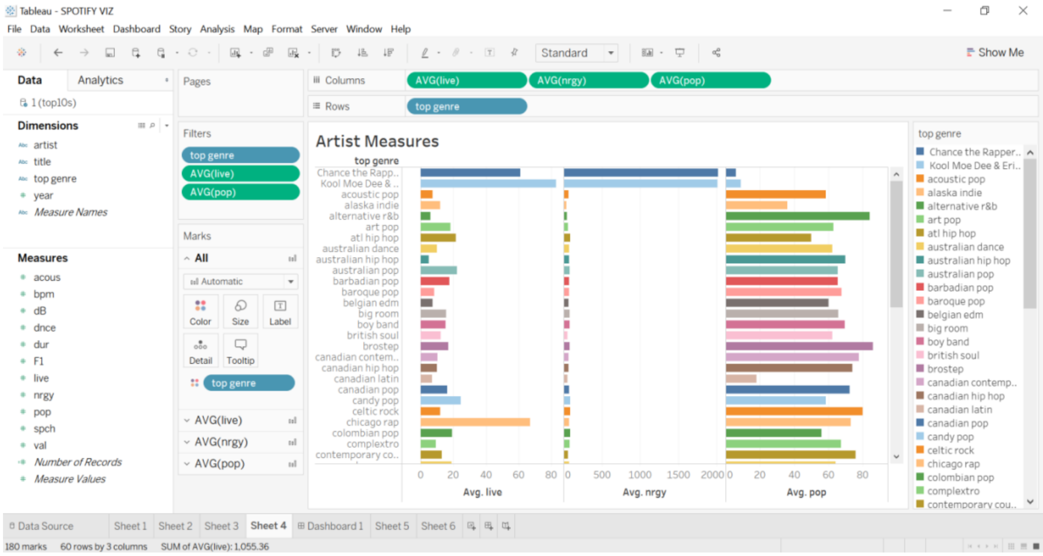
Dimensions and measures are added to the ‘Columns’ and ‘Row’ section. Only year attribute in column and acousticness, popness, speechiness is taken as row attribute. ‘AVG()’ signifies the operation. Here we can see ‘path’ is used to visualize data. A pattern can be spotted for average acousticness, as it mostly increases and decreases alternatively at constant intervals. Similarly for speechness. For popness however, data is completely constant per year. From this we can infer that around 2012, the rise of EDM music was seen where most of the DJ’s and artists started creating more upbeat non-acoustic dance music with lesser lyrics in the songs.



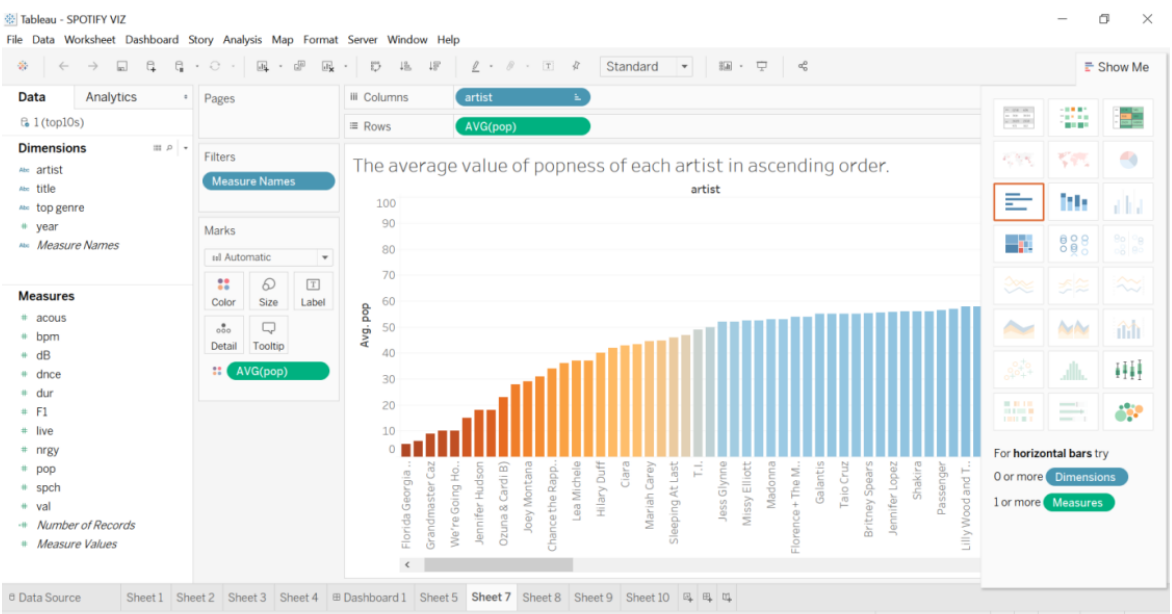
Filters used here are as per year. Row attributes here include number of records and column attribute is year. As you can see the ‘SUM()’ operation is used. All the records are summed up for all year separately. As seen from the graph some of the circles are of similar size, such as 2013,2017 and 2018 here is where color key plays an important role on the left-hand side. This instantly shows that 2013<2017<2018 also giving us an estimation of the exactly quantity. 2015 proves to be a clear winner based on the number of records while 2019 lags behind.



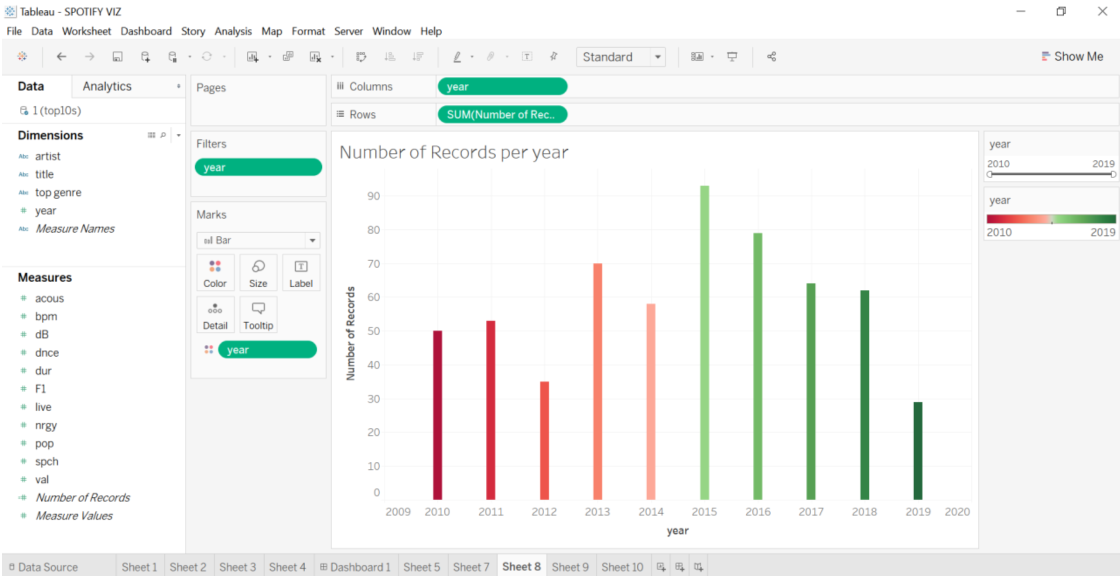
Filters used here is measure names. Row attributes here include artists and measure names. All the records are summed up for all year separately. This visualization is a color map. On the left-hand side you can see that all the attributes such as acoustics, pop ,speechness and value were considered and each is marked as a different colour. All the attributes are compared with each other under the category of an artist as given here in the graph. The motive of this is to find a correlation between all the variables. Here we also have a highlighter which helps us to search for a particular artist for which we need to see the details of the measures.



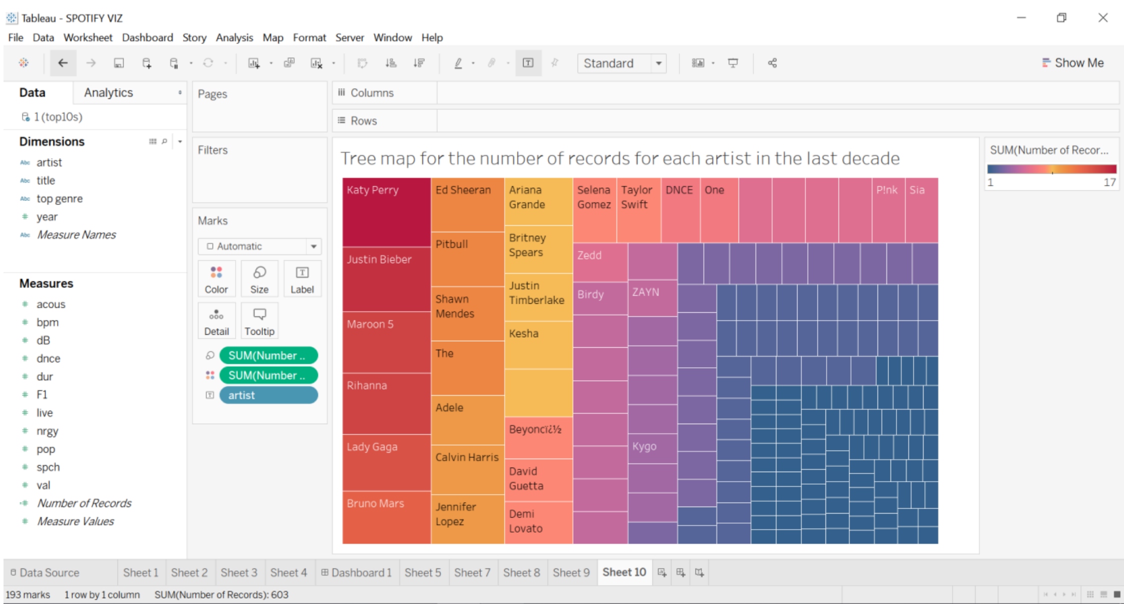
Filters used here is top genre. Row attributes here include artists and measure names. All the records are summed up for all year separately. This visualization is a color map. On the left-hand side you can see that all the attributes such as acoustics, pop ,speechness and value were considered and each is marked as a different colour. All the attributes are compared with each other under the category of an artist as given here in the graph. The motive of this is to find a correlation between all the variables.



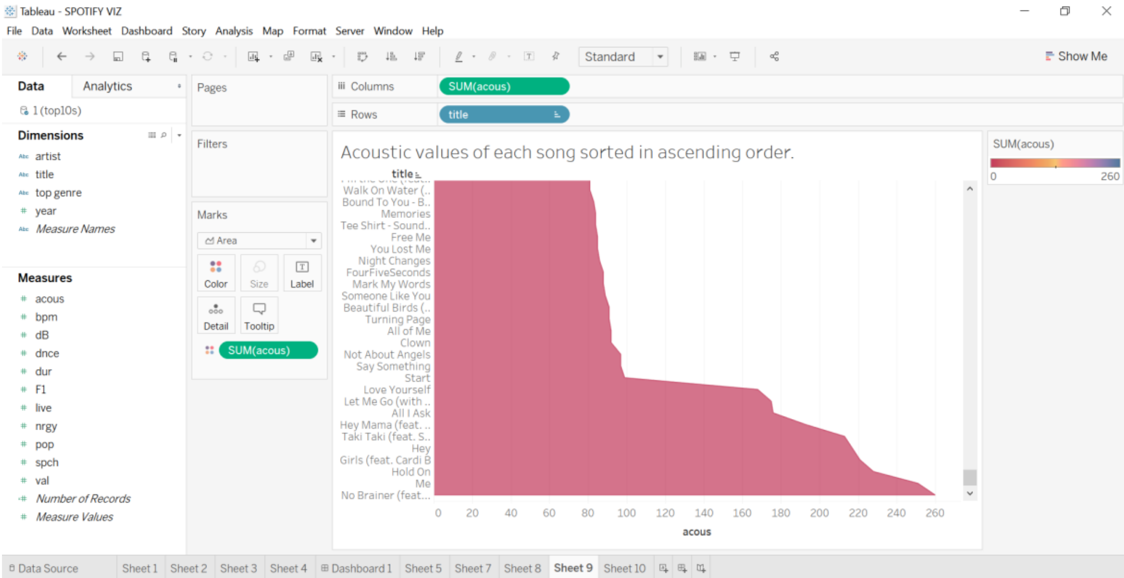
Filters used here is measure names. Row attributes here include average of popness. The average of popness in numerical value is defined by the color pattern bar. All the records are summed up for all year separately. This visualization is a color map bar graph. The purpose of this map is to find the overall popness score each of the artists that gives us an idea of their genre. Here the row is the average pop and the column is the artist.



The above visualization is a bar graph. Here two colour pattern bars are given two represent the year. The row is the sum of records and the column is the year. Here the year filter is used so that while visualizing the graph we can put certain constraints on the years that we must view.



The above visualization is a tree map for the number of records for each artist in the last decade. If we want to know which artist released the maximum number of songs in the decade we can look at this tree graph by a quick glance we can say katy perry has the highest number of albums in the decade. No filters can be used here. Size and color were representing the sum of number of records for each artist.



The software tableau has a special ‘sort’ feature. Using this feature we can select our attribute in this case we have chosen acoustics. This feature allows us to sort the results of the attribute in any fashion that we want and some of the options include:

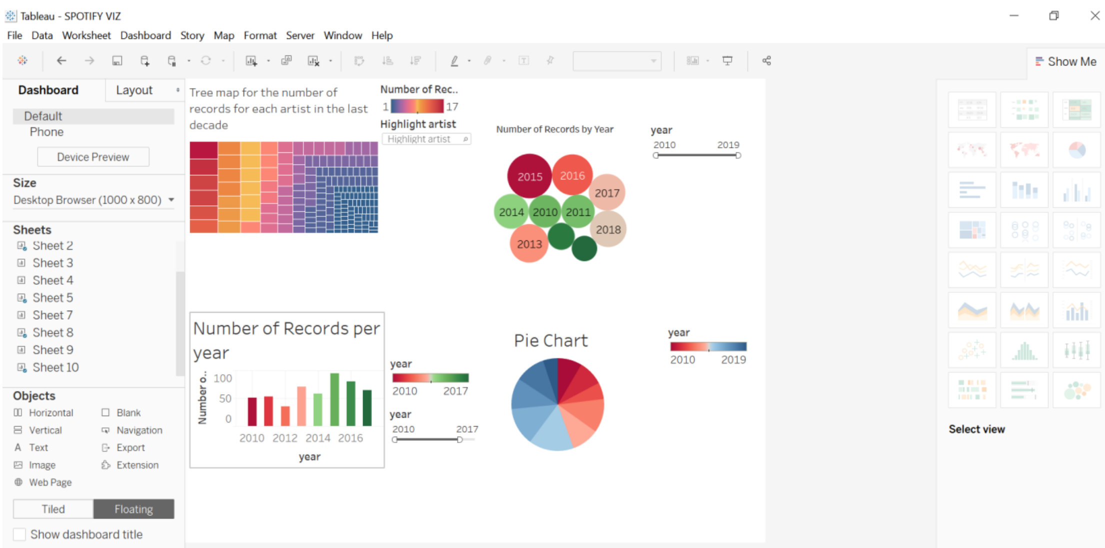
- Alphabetically
- Increasing
- Decreasing

One pattern colour bar given is basically representing the sum of acoustics. No filters are used here. Columns show the sum of acoustics and the rows shows the title.

Creation and Analysis of the Dashboard

Tableau is a very user friendly tool in which we can make dashboards very easily.

After we have created different sheets we can add them on a single dashboard. A dashboard is a collection of several views, letting you compare a variety of data simultaneously. Here we have included all the different forms in which we can display the Number of records per artist as well as by year.



Using K-Means Clustering for making recommendation playlists

The above analysis of the dataset can be used in making playlists of similar songs using machine learning techniques such as K-Means Clustering. Here we have used scikit-learn library of python for our implementation. Here we have created 3 playlists using K-Means clustering based on the 'popness' and 'acoustic' values of songs. These clusters automatically give us songs of similar types which can be placed in a single playlist. These recommended playlists ensure that people are more likely to listen to songs which are similar in vibe.

```
#The necessary packages are imported.

from sklearn.cluster import KMeans
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
%matplotlib inline
```

Data Pre-processing:

First, we imported our dataset as a csv file. Then we made our target dataset smaller by extracting all the songs from the year 2016. We have scaled the 'pop' and 'acous' values between 0 and 1 using MinMaxScaler to get a more accurate result in our algorithm.

```
#Here we import our dataset as a csv file.
#We have extracted all the songs from the year 2016.

df = pd.read_csv("Downloads/top10s.csv")
df = df[df.year==2016]
df.head()
```

	Unnamed: 0	title	artist	top genre	year	bpm	nrgy	dnce	dB	live	val	dur	acous	spch	pop
363	364	The Hills	The Weeknd	canadian contemporary r&b	2016	113	56	58	-7	14	14	242	7	5	84
364	365	Love Yourself	Justin Bieber	canadian pop	2016	100	38	61	-10	28	52	234	84	44	83
365	366	Cake By The Ocean	DNCE	dance pop	2016	119	75	77	-5	4	90	219	15	5	81
366	367	Don't Let Me Down	The Chainsmokers	electropop	2016	160	87	53	-5	14	42	208	16	17	81
367	368	In the Name of Love	Martin Garrix	big room	2016	134	52	50	-6	45	17	196	11	4	81

```python
#Here we scale our values for an efficient system.

sc = MinMaxScaler()
df['pop'] = sc.fit_transform(df[['pop']])
df['acous'] = sc.fit_transform(df[['acous']])
df[['pop','acous']]
```

	pop	acous
363	1.000000	0.070707
364	0.988095	0.848485
365	0.964286	0.151515
366	0.964286	0.161616
367	0.964286	0.111111
368	0.952381	0.020202
369	0.952381	0.202020

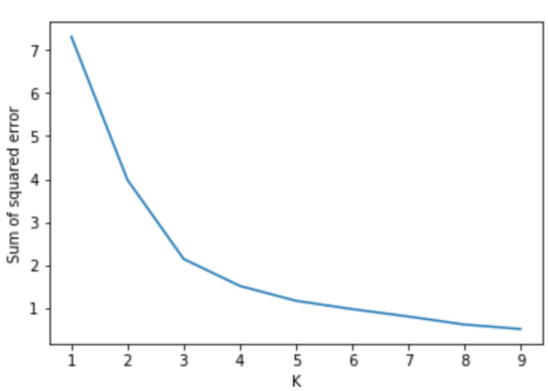
Evaluating the clusters using the Elbow Method:

```python
#Here we find the Sum of Squared Error value for each value of k

k_rng = range(1,10)
sse = []
for k in k_rng:
    km = KMeans(n_clusters=k)
    km.fit(df[['pop','acous']])
    sse.append(km.inertia_)
sse
```

```
[7.308102474901041,
 3.9803750657789556,
 2.1442709346572597,
 1.5186851291378707,
 1.1731579774509067,
 0.9799718469440778,
 0.80813799018053811,
 0.62097598513049801,
 0.5180501938569434]
```

```python
#We have plotted the sse values to see that at k=3 clusters, we get the optimal solution.

plt.xlabel('K')
plt.ylabel('Sum of squared error')
plt.plot(k_rng,sse)
```

```
[<matplotlib.lines.Line2D at 0x1ef5f8d6080>]
```



Applying the K-Means Algorithm:

```
#Here we use K-Means clustering algorithm on the dataset.

km = KMeans(n_clusters=3)
km
```

```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
    n_clusters=3, n_init=10, n_jobs=1, precompute_distances='auto',
    random_state=None, tol=0.0001, verbose=0)
```

```
y_predicted = km.fit_predict(df[['pop','acous']])
y_predicted
```

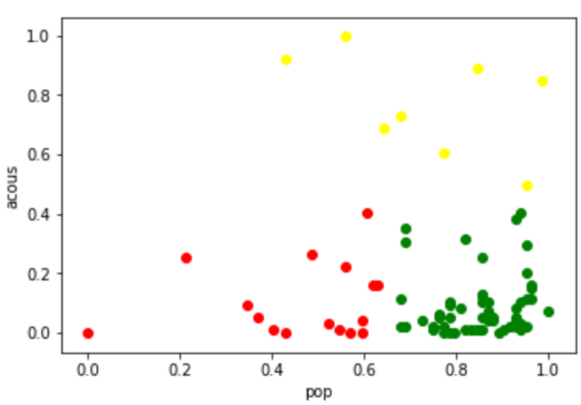
```
array([0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 2, 2, 2, 2, 2, 2, 1,
       2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2])
```

```
df['cluster'] = y_predicted
df.head()
```

	Unnamed: 0	title	artist	top genre	year	bpm	nrgy	dnce	dB	live	val	dur	acous	spch	pop	cluster
363	364	The Hills	The Weeknd	canadian contemporary r&b	2016	113	56	58	-7	14	14	242	0.070707	5	1.000000	0
364	365	Love Yourself	Justin Bieber	canadian pop	2016	100	38	61	-10	28	52	234	0.848485	44	0.988095	1
365	366	Cake By The Ocean	DNCE	dance pop	2016	119	75	77	-5	4	90	219	0.151515	5	0.964286	0
366	367	Don't Let Me Down	The Chainsmokers	electropop	2016	160	87	53	-5	14	42	208	0.161616	17	0.964286	0
367	368	In the Name of Love	Martin Garrix	big room	2016	134	52	50	-6	45	17	196	0.111111	4	0.964286	0

```
df0=df[df.cluster==0]
df1=df[df.cluster==1]
df2=df[df.cluster==2]

plt.scatter(df0['pop'],df0['acous'],color='green')
plt.scatter(df1['pop'],df1['acous'],color='yellow')
plt.scatter(df2['pop'],df2['acous'],color='red')

plt.xlabel('pop')
plt.ylabel('acous')
```

```
Text(0,0.5,'acous')
```



Displaying the clusters as playlists:

```
#This is one of the clusters which can be a playlist.

df2[['title','pop','acous']]
```

	title	pop	acous
425	Make Me... (feat. G-Eazy)	0.630952	0.161616
426	Keeping Your Head Up	0.619048	0.161616
427	True Colors	0.607143	0.404040
428	Make Me Like You	0.595238	0.040404
429	Champagne Problems	0.595238	0.000000
430	Blown	0.571429	0.000000
432	Pep Rally	0.559524	0.222222
433	Higher	0.547619	0.010101
434	Invitation	0.523810	0.030303
435	One Call Away (feat. Tyga) - Remix	0.488095	0.262626
437	Little Lies	0.428571	0.000000
438	Do You Wanna Come Over?	0.404762	0.010101
439	BURNITUP!	0.369048	0.050505
440	Picky - Remix	0.345238	0.090909
441	Behind Your Back	0.214286	0.252525
442	Million Years Ago	0.000000	0.000000

Conclusion

Spotify is the largest music streaming service available. The company started in 2006 in a time when piracy caused considerable losses to the music industry. In January 2015 they had 60 million users in total of which 15 million premium users (1) and these numbers seem to be increasing. Spotify offers free streaming of music to its users, though one can purchase a premium membership for added benefits, such as no advertisements and being able to listen to music offline. The large number of users and content of Spotify create a large database of users and songs that users listened to that could hold interesting patterns and information for related companies, such as Spotify themselves, record companies or radio stations. The dataset in question has been provided by, so we first look for general applications of the data and then focus on possibilities that will also be useful to , but will also be interesting from a scientific perspective. By performing some statistics on the entire dataset, we try to determine the worth of the Spotify data for both and for scientific purposes. We will answer a few relatively simple questions regarding interesting patterns found in the data and try to formulate a good model that can be used with this data. We used the software tableau to visualize the data. We created many different types of visualizations that included:

- Bubble cluster
- Bar graphs
- Tree map
- Path graph

Each different visualization had different advantages and disadvantages. Not only that, we visualized data considering many different combination of attribute factors. As tableau is a very user friendly software it is widely used by companies and in business analysis.

Tableau Dashboards provide a wholesome view of your data by the means of visualizations, visual objects, text, etc. Dashboards are very informative as they can present data in the form of stories, enable the addition of multiple views and objects, provide a variety of layouts and formats, enable the users to deploy suitable filters. You even have the option to copy a dashboard or its specific elements from one workbook to another easily.

Advanced Visualizations (Chart Types)

One of the key features of Tableau and the one that got its popularity is its wide range of visualizations. In Tableau, you can make visualizations as basic as a:

- Bar chart
- Pie chart

and as advanced as a:

- Histogram
- Gantt chart
- Bullet chart
- Motion chart
- Treemap
- Boxplot

and many more. You can select and **create any kind of visualization** easily by selecting the visualization type from the Show Me tab.

How to use these visualizations?

- Marketing strategies can be created based on the most popular artist and genres that are at the top. The company may ask themselves "which artist sells?". Online advertisements can be created accordingly.
- Spotify can organize its target user-base as the time period and type of music that people like to listen to at that particular moment.
- Such a visualization and interpretation as well as data collection is of importance. Data is the most valuable asset in today's day and age such data can be sold to music companies for large sum of money.

Throughout the course of our project, we learned a number of things about working with large datasets and how to properly analyze them. One of the biggest lessons is how inconvenient data rate limiting can be. We also learned that data cleaning is a much more important step in the process of retrieving a dataset than we first anticipated. Throughout multiple steps in the process, there were many extra bits of unnecessary or redundant information that needed to be filtered out. Had we not caught these redundancies, it is likely our final analysis would be dramatically skewed. It was particularly important to make sure this step was done correctly in order to make sure results could actually be pulled from the Spotify search.

Some key findings from the graphing stage is the discovery of most common keys and tempos. It seems likely that all keys and tempos would have had a roughly equal representation, but instead we found that there were some groups that were much more likely than others. Interestingly, the distribution of keys followed notes from the C-major scale. Tempos around 100 bpm were also by far the most common.

We also discovered that changes in the energy and acousticness of songs, combined with decreased correlations between tempo and valence, support preconceived notions that in recent years, 4-beat rock-n-roll style songs, which show heavy correlations between tempo and valence, are being increasingly pushed out of the mainstream and replaced by electronically produced songs of varying tempos.

We also discovered that there was a transition of the kind of music over the years which included songs to be more upbeat and more EDM songs were getting popular over the later half of the decade. This shows that the music taste of the people is a constant changing process. Artists like Katy Perry, Maroon 5 and Justin Beiber were the major players in the game. The produced the most number of songs with a high success rate. More and more Dj's like Martin Garrix, David Guetta and Alan Walker were also becoming main stream and more popular. One of the major takeaways was also that the music industry was at its peak during 2015-2016 in terms of quantity as well as the quality of songs.

The major application of all this analysis can be put in use for the machine learning algorithms of the Spotify API to create recommendation playlists for the users by mixing similar valued songs to each playlist to make it even more attractive and enjoyable for the users.

References

1. Analyzing the Spotify Top 200 Through a Point Process Lens Michelangelo Harris, Brian Liu, Cean Park, Ravi Ramireddy, Gloria Ren, Max Ren, Shangdi Yu, Andrew Daw, and Jamol Pender School of Operations Research & Information Engineering Cornell University, Ithaca, NY October 4, 2019

2. Approaching Media Industries Comparatively: A Case Study of Streaming, Daniel Herbert, Amanda Lotz and Lee Marshall

3. Spotify Teardown: Inside the Black Box of Streaming Music by Maria Eriksson, Rasmus Fleischer, Anna Johansson, Pelle Snickars & Patrick Vonderau, The MIT Press, ISBN 978-0262038904

4. Infinite content and interrupted listening, The impact of smartphones, streaming and music ‘superabundance’ on everyday personal music listening behaviour, Ellen Moore, July 2019

5. Metrics and decision-making in music streaming, Arnt Maasa and Anja Nylund Hagen

6. Music in Streams: Communicating Music in the Streaming Paradigm, Anja Nylund Hagen, Postdoctoral fellow, Dept. of Musicology, University of Oslo

7. Music Radio as a Format Remediated for Stream-Based Music Use, Andreas Lenander AEgidius

8. Music Through the Ages: Trends in Musical Engagement and Preferences From Adolescence Through Middle Adulthood, Article in Journal of Personality and Social Psychology · July 2013, Arielle Bonneville-Roussy, Peter J. Rentfrow, and Man K. Xu University of Cambridge and Jeff Potter Atof Inc., Cambridge, Massachusetts

9. Musical trends and predictability of success in contemporary songs in and out of the top charts, Interiano M, Kazemi K, Wang L, Yang J, Yu Z, Komarova NL. 2018 Musical trends and predictability of success in contemporary songs in and out of the top charts.R. Soc. open sci.5: 171274.

10. Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery, Hannes Datta George Knox Bart J. Bronnenberg* Forthcoming at Marketing Science

11. The rise of streaming music and implications for music production, R. Scott Hiller and Jason Walter

12. Temporal Trends in Music Popularity A quantitative analysis of Spotify API data, Matthew Menten, Kieren Ng, Braden Holmes and Tanner O’Rourke

13. THE DAILY USE OF MUSIC STREAMING PLATFORMS BY THE DUTCH MILLENNIAL AGE GROUP ORIGINALLY EXECUTED FOR XITE NETWORKS B.V. RAMESH SITAL, BBA MA Applied Musicology Faculty of Humanities Utrecht University, 2019

14. ORIENTALISKA STUDIER, Special Issue International Conference: 60th Anniversary of Diplomatic Relations between the Kingdom of Sweden and the Republic of Korea (ROK), May 23-24, 2019

15. THE PAST AND FUTURE OF MUSIC LISTENING Between Freeform DJs and Recommendation Algorithms,Elena Raz/ogova

16. www.spotify.com

17. https://www.tableau.com/about/blog/2019/7/how-visualize-spotify-music-trends-tableau

18. www.tableau.com

19. https://scikit-learn.org/

20. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

21. https://en.wikipedia.org/wiki/Spotify

22. https://www.kaggle.com/datasets?search=spotify

23. https://www.tutorialspoint.com/tableau/tableau_dashboard.htm

24. https://www.pocket-lint.com/apps/news/spotify/139236-what-is-spotify-and-how-does-it-work

25. https://towardsdatascience.com/building-music-playlists-recommendation-system-564a3e63ef64