RESOURCE

Analysing Spotify Data

**Learning Level:** First (P2 to P4), Second (P5 to P7)

**PPDAC Framework:** Data|Analysis|Conclusion

**Theme:**  Data & Analysis | Searching, processing and managing information

responsibly

**Curricular Area:** Maths & Numeracy, Technologies

**Duration:** 1 hour

**Materials:**

• Spotify dataset appropriate to your learners. [Top 20](https://dataschools.education/wp-content/uploads/2020/08/Top20SpotifySongsSimplifiedData.xlsx) or [Top 50](https://dataschools.education/wp-content/uploads/2020/08/Top50SpotifySongs2019.xlsx).

• Learner worksheet that corresponds to the data set in use

• Paper, pencil, ruler

• Coloured paper & coloured pencils (optional)

• Access to spreadsheet software (optional)

Icon

Description automatically generated

Icon

Description automatically generatedA picture containing clock

Description automatically generated

# Introduction

This task makes use of secondary data provided by Spotify. Learners will use the data to identify patterns and relationships between different characteristics. The activity will support learners in developing their ability to review and interpret a dataset. The activity starts by encouraging learners to think of questions that the dataset might answer and thus develop their “Problem” skills. Once the dataset has been analysed learners will have the opportunity to reorganise and restructure data to help them answer their questions (draw conclusions)

# Curriculum Links

This activity can be completed with or without a computer. The curriculum links for both options are shown below.

**Curriculum Links for Offline Task**

* **MTH 1-21a:** Using technology and other methods, I can display data simply, clearly and accurately by creating tables, charts and diagrams, using simple labelling and scale.
* **MNU 1-20a:** I have explored a variety of ways in which data is presented and can ask and answer questions about the information it contains.
* **MNU 2-20a:** Having discussed the variety of ways and range of media used to present data, I can interpret and draw conclusions from the information displayed, recognising that the presentation may be misleading.
* **MTH 2-21a:** I can display data in a clear way using a suitable scale, by choosing appropriately from an extended range of tables, charts, diagrams and graphs, making effective use of technology**.**

**Curriculum Links for Online Task**

* **TCH 1-02a:** Using digital technologies responsibly I can access, retrieve and use information to support, enrich or extend learning in different contexts.
* **TCH 2-01a**: I can extend and enhance my knowledge of digital technologies to collect, analyse ideas, relevant information and organise these in an appropriate way.
* **TCH 1-13a**: I can explore and comment on processes in the world around me making use of core computational thinking concepts and can organise information in a logical way.

# Differentiation

This task can be tailored between first and second level by varying the dataset used.

# Pre-Requisites

This activity can be used to introduce learners to elements of PPDAC or to provide a further opportunity to develop their understanding.

# Discussion

Hold a discussion with the group about music. Start with favourite songs and artist and support the group to share their own favourites. Progress to asking what makes a good song and why it might be popular. Finally start to identify characteristics of a song and note them either on a board or encourage the learners to write them on a piece of paper. Characteristics can include.

* Song name
* Artist name
* Genre/Type
* Length
* Rank (popularity)

# Exploring the dataset (suitable for home)

This step is focused on the “**Problem**” step of PPDAC. We are encouraging learners to think about interesting questions about the data they have.

Look through the printable ‘[Top 20’](https://dataschools.education/wp-content/uploads/2020/08/Top20SpotifySongsSimplifiedData.xlsx) or ‘[Top 50’](https://dataschools.education/wp-content/uploads/2020/08/Top50SpotifySongs2019.xlsx) Spotify songs on the ‘Spotify Analysis’ resource page of the Data Education in Schools website or at the end of this document, or access the data via Kaggle at [www.kaggle.com](http://www.kaggle.com).

What do you notice about the data, what can you find out? Can you think of any questions you would like to know the answer to?

Write down:

* 3 different genres of music in your dataset
* 2 songs with the same popularity score
* The song that will come first and last alphabetically

# Organising & Analysing the dataset

This task if focused on the **“Analysis”** step of PPDAC. We are encouraging learners to review the dataset and identify useful information that may help them understand the questions they thought about in step 1.

If completing this task offline then it can be easier to use the smaller dataset (20 songs) even if the learners are at second level. The task is split into two, one set of questions for each dataset. Worksheets are available for each dataset at the end of this document.

**Smaller Dataset**

Use your dataset to complete the following tasks.

1. What is the most popular song?
2. What is the least popular song?
3. Group the songs by genre
   1. How many songs are there in the pop category?
   2. What categories only have 1 song in them?
4. Rewrite the table starting at the most popular song and ending with the least popular song.

**Larger Dataset**

Use your dataset to complete the following tasks.

1. What is the most popular song?
2. What is the least popular song?
3. Group the songs by genre
   1. How many songs are there in the pop category?
   2. What categories only have 1 song in them?
4. Which song is the best to dance to?
5. Which artist sings the song with the most beats per minute?
6. Identify 3 artists with more than one song in the dataset

The above tasks can be made easier if the data is sorted in Excel (or similar spreadsheet package). Teachers or Parents can encourage children to learn how to sort the data by following the instructions here: <https://youtu.be/KS9N4yAjuYQ>

# Visualise your dataset

This task mixes the **“Analysis”** and **“Conclusions”** part of PPDAC (Problem, Plan, Data, Analysis, Conclusion) cycle.Creating a summary table continues the analysis phase and seeks to show how a large amount of data can be presented in a different way. Visualising the data can support both analysis and drawing conclusions. Learners should be encouraged to connect questions they asked at the start to the analysis they are doing the information they can extract from the visualisation.

We are now going to learn to create a visual representation of the data. Again this can be done with or without a computer.

**Smaller Dataset**

1. Add up the total score for each genre of music. Complete this table.

|  |  |
| --- | --- |
| **Genre** | **Total Score** |
| Canadian hip hop |  |
| Canadian pop |  |
| country rap |  |
| dance pop |  |
| dfw rap |  |
| electropop |  |
| escape room |  |
| Latin |  |
| Panamanian pop |  |
| pop |  |
| reggaeton |  |
| reggaeton flow |  |
| trap music |  |

1. Create a bar chart of genre and total score. The chart can be created in a variety of different ways.
   1. Draw it on paper
   2. Use Lego, or similar blocks to create it on a table
   3. Use sticks and other natural materials to create it outside
   4. Use Excel to create a chart in your spreadsheet.

The group can have an optional discussion about whether the highest scoring genre is therefore the most popular. This is an opportunity consider the ideas behind averages.

Learners could be asked to:

* Review what are the 3 most popular songs, are these in the most popular genre?
* What might cause the most popular genre to be different to the genre of the most popular song?

**Larger Dataset**

The following tasks are more suited to using a spreadsheet.

1. Create a table with 4 columns, Genre, Number of songs, Total Popularity Score and Average score
2. Fill in your table
   1. Put each genre in the first column
   2. Find out how many songs are in each genre and enter it in column 2
   3. Add up total popularity score for each genre and enter it in column 3
3. Calculate the average score by dividing the total popularity score by the number of songs and enter it in column 4.
4. Create a bar chart of genre and average score in your spreadsheet software.

# Dataset Top 20

[Excel version of Top 20 dataset](https://dataschools.education/wp-content/uploads/2020/08/Top20SpotifySongsSimplifiedData.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Track.Name | Artist | Genre | Popularity |
| 15 | Money In The Grave (Drake ft. Rick Ross) | Drake | canadian hip hop | 92 |
| 19 | Lalala | Y2K | canadian hip hop | 88 |
| 1 | Señorita | Shawn Mendes | canadian pop | 79 |
| 9 | Old Town Road - Remix | Lil Nas X | country rap | 87 |
| 12 | Loco Contigo (feat. J. Balvin & Tyga) | DJ Snake | dance pop | 86 |
| 3 | boyfriend (with Social House) | Ariana Grande | dance pop | 85 |
| 16 | No Guidance (feat. Drake) | Chris Brown | dance pop | 82 |
| 5 | Goodbyes (Feat. Young Thug) | Post Malone | dfw rap | 94 |
| 18 | Sunflower - Spider-Man: Into the Spider-Verse | Post Malone | dfw rap | 91 |
| 10 | bad guy | Billie Eilish | electropop | 95 |
| 20 | Truth Hurts | Lizzo | escape room | 91 |
| 17 | LA CANCIÓN | J Balvin | latin | 90 |
| 14 | Otro Trago - Remix | Sech | panamanian pop | 87 |
| 8 | How Do You Sleep? | Sam Smith | pop | 90 |
| 13 | Someone You Loved | Lewis Capaldi | pop | 88 |
| 4 | Beautiful People (feat. Khalid) | Ed Sheeran | pop | 86 |
| 6 | I Don't Care (with Justin Bieber) | Ed Sheeran | pop | 84 |
| 11 | Callaita | Bad Bunny | reggaeton | 93 |
| 2 | China | Anuel AA | reggaeton flow | 92 |
| 7 | Ransom | Lil Tecca | trap music | 92 |

# Dataset Top 50

[Excel version of Top 50 dataset](https://dataschools.education/wp-content/uploads/2020/08/Top50SpotifySongs2019.xlsx)

[PDF printable](https://dataschools.education/wp-content/uploads/2020/08/Top-50-Spotify-Songs-Printable.pdf)

A picture containing table, calendar

Description automatically generatedA picture containing table

Description automatically generatedA picture containing table, calendar

Description automatically generated

Learner Worksheet – Top 20 Songs

# Exploring the dataset

Look at the table of songs.

What kind of information do you see in the data?

Can you think of any questions that you might ask about the songs Write down:

* 3 different genres of music in your dataset
* 2 songs with the same popularity score
* The song that will come first and last alphabetically
  + First alphabetically:
  + Last alphabetically:

# Organising & Analysing the dataset

Use your dataset to complete the following tasks.

1. What is the most popular song?
2. What is the least popular song?
3. Group the songs by genre
   1. How many songs are there in the pop category?
   2. What categories only have 1 song in them?
4. Rewrite the table starting at the most popular song and ending with the least popular song. You can do this on the back of this page.

# Visualise your dataset (without a computer)

We are now going to learn to create a visual representation of the data.

1. Add up the total score for each genre of music. Complete this table.

|  |  |
| --- | --- |
| **Genre** | **Total Score** |
| canadian hip hop |  |
| canadian pop |  |
| country rap |  |
| dance pop |  |
| dfw rap |  |
| electropop |  |
| escape room |  |
| latin |  |
| panamanian pop |  |
| pop |  |
| reggaeton |  |
| reggaeton flow |  |
| trap music |  |

1. Create a bar chart of genre and total score. You can create your chart in different ways.
   1. Draw it on paper
   2. Use Lego, or similar blocks to create it on a table
   3. Use sticks and other natural materials to create it outside
   4. Use excel to create a chart in your spreadsheet.
2. From your chart, which is the most popular genre?
3. Think about the results. Do you agree with them or do you think another genre is more popular? Write down what you think.

Learner Worksheet – Top 50 Songs

# Exploring the dataset

Look at the table of songs. What kind of information do you see in the data?

Can you think of any questions that you might ask about the songs. Write down:

* 3 different genres of music in your dataset
* 2 songs with the same popularity score
* The song that will come first and last alphabetically
  + First alphabetically:
  + Last alphabetically:

# Organising & Analysing the dataset

Use your dataset to complete the following tasks.

1. What is the most popular song?
2. What is the least popular song?
3. Group the songs by genre
   1. How many songs are there in the pop category?
   2. What categories only have 1 song in them?
4. Which song is best to dance to?
5. Which artist sings the song with the most beats per minute?
6. Identify 3 artists with more than one song in the dataset.

# Visualise your dataset (with a computer)

Use your spreadsheet to visualise your dataset.

1. Create a new tab
2. Rename your tab “Summary Data”
3. Create a new table with 4 columns, Genre, number of songs, Total Popularity Score and average score
4. Fill in your table
   1. Put each genre in the first column
   2. Find out how many songs are in each genre and enter it in column 2
   3. Add up total popularity score for each genre and enter it in column 3
   4. Calculate the average score by dividing the total popularity score by the number of songs and enter it in column 4
5. Create a bar chart of genre and average score in your spreadsheet software
6. From your chart, which is the most popular genre?
7. Think about the results. Do you agree with them or do you think another genre is more popular? Write down what you think.

1. Pick any two types of information and create a new table or chart.
2. Write down something interesting you have discovered in your data.

**If you require this document in an alternative format, such as large print or a coloured background, please contact Claire Sowton,** [**data.schools@ed.ac.uk**](mailto:data.schools@ed.ac.uk) **or Moray House School of Education and Sport, St John’s Land, Holyrood Road, Edinburgh, EH8 8AQ**

A close up of a sign

Description automatically generatedA picture containing clock

Description automatically generated