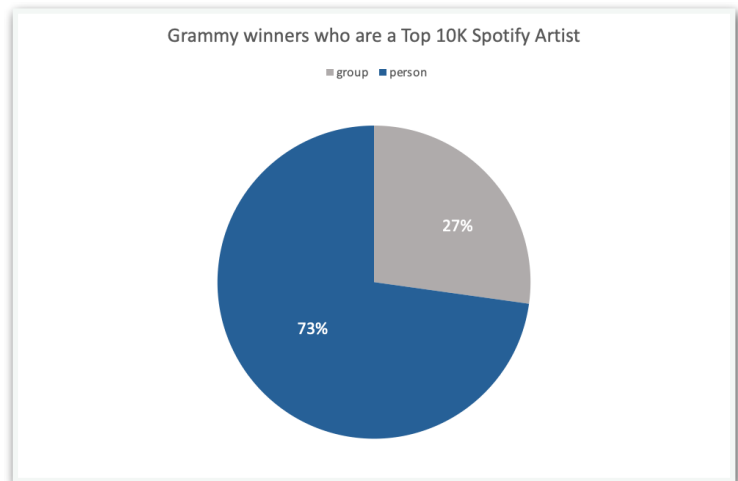


i) Project Summary

This project includes the cleaning and wrangling of a dataset of Grammy awards for artists and their various attributes such as categories. I also used the top 10,000 artists on Spotify and other information such as their gender and age. Using these two datasets, I wished to answer four questions regarding the Grammys and Spotify, which will help me to infer whether the popularity of Spotify artists necessarily translates into getting Grammy awards in the music industry. I was able to obtain all the answers through various wrangling skills in software such as Excel and MongoDB.

1) How many groups have won at least one Grammy and also are in the top 10K Spotify artists?

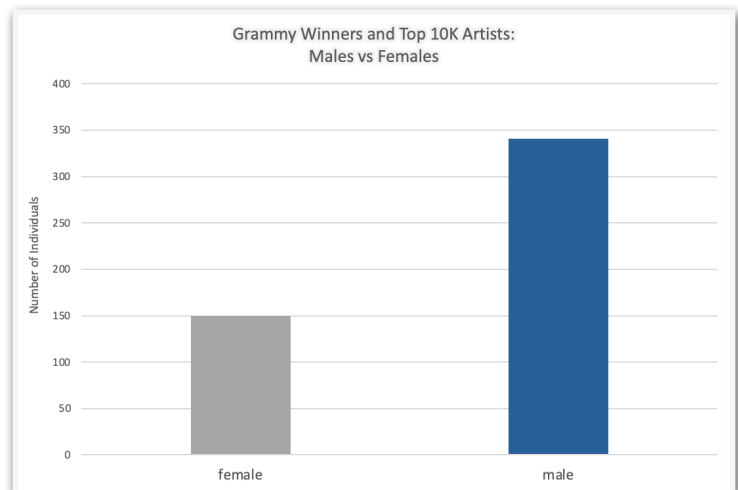
I deduced 53 unique groups (not individuals) that have at least one Grammy that are also in the Top 10,000 Spotify Artists list. Without counting distinct groups, in general, group winners make up about 27% of the awards given. This can be seen by the chart to the right:



Appendix 1

2) Of those who are a Top 10K artist on Spotify, have females won more Grammy awards than males?

I found through my analysis that the answer is no! In fact, males have won about 191 more Grammy awards than females.



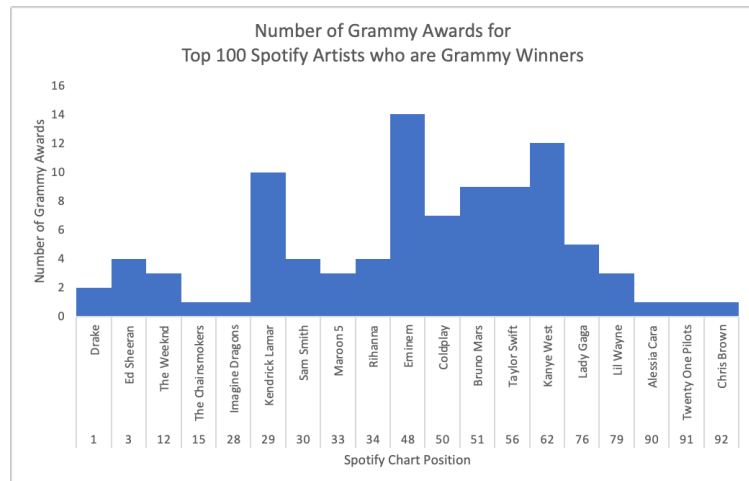
Appendix 2

3) Which Grammy Award category has the most winners equal to or under the age of 50?

I have concluded that 'Best Rap Album' has the most Grammy winners equal to or under the age of 50, with 15 awards by artists won in this category.

4) Of the Top 100 Artists in Spotify's Top 10K (who have won Grammy awards), is their number of Grammy awards higher depending on their popularity on Spotify?

I have deduced from the analysis of the chart I have created below, that just because an artist is more popular on Spotify (in the Top 100 etc.) does not necessarily mean that they get a higher amount of Grammy awards. This was done by counting the number of Grammys that each artist in the Top 100 have received. However, being in the Top 100 does seem to provide a very good chance of receiving any award at all.

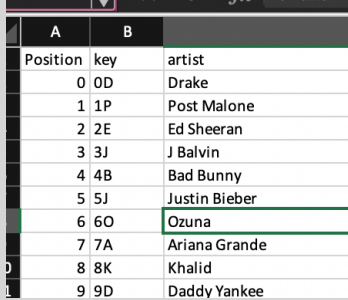

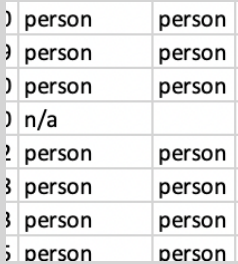
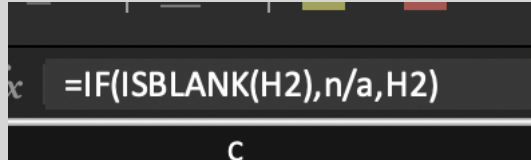
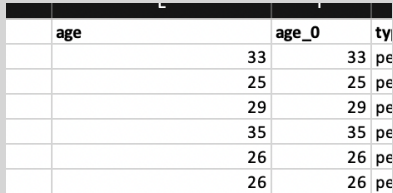
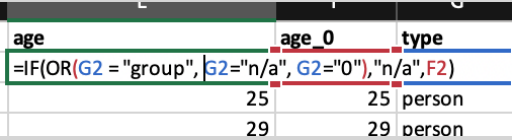
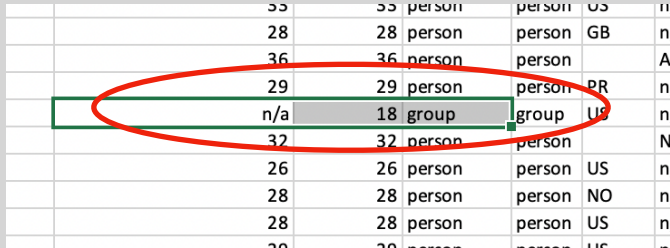


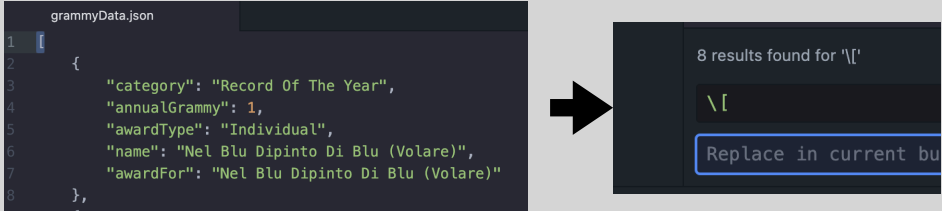

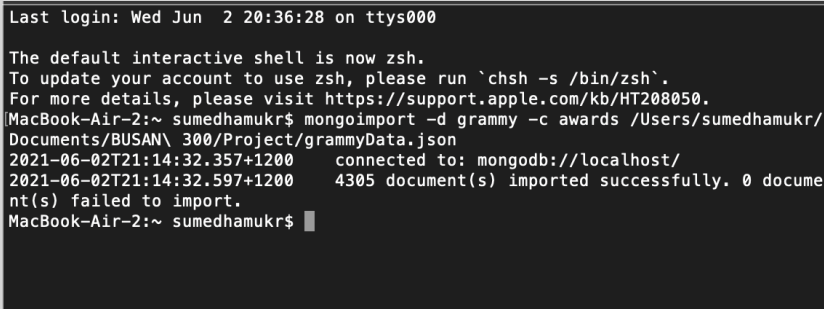
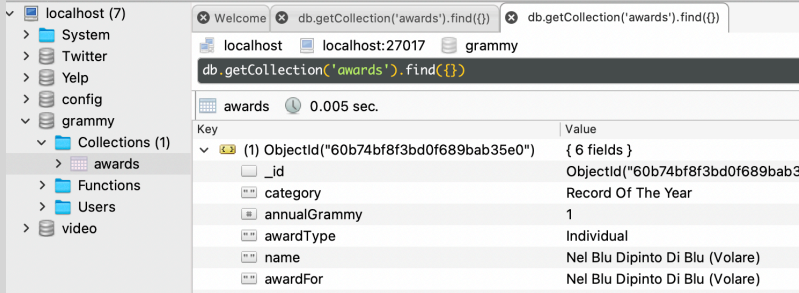
Appendix 3

ii) Wrangling Details

Date	Task	Procedure
28/4	Sourcing datasets	<p>Link to the Top 10K Spotify Artists: https://www.kaggle.com/jackharding/spotify-artist-metadata-top-10k?rvi=1</p> <p>Link to the Grammy Awards Data: https://www.kaggle.com/theriley106/grammyawardsinnumbers</p>
14/5	Analysing formats / structure	<p>Spotify artists data: The data is in .csv format, and after inspection, there are 12 columns and 10001 rows (including the column titles). The first column appears to be the artist's position in the top10K, starting from 0.</p> <p>Grammy awards data: The data is in JSON format and after inspecting it, there are 5 attributes. (name, category, annualGrammy, awardType, awardFor).</p>
29/5	Analysing potential problems with datasets	<p>Top 10K spotify artists csv data: I imported from text with comma delimited columns, and no text qualifier.</p> <p>Validity:</p> <ul style="list-style-type: none"> - Every artists' name is mentioned (no blanks, n/a's). Correct format and structure. - Index column could be used as unique identifier (however the index starts from 0 again at artist 3166). <p>Completeness:</p> <ul style="list-style-type: none"> - After importing in Excel, many 'n/a's in various columns (especially the district columns) and blanks. - Age column: many ages are currently '0' for a lot of artists. <p>Accuracy:</p> <ul style="list-style-type: none"> - For ages of groups, the ages are set to random numbers, I believe. This may need to be changed to something like 'multiple', or blank. - Redundancy of district_1 columns (they provide little to no information). <p>Grammy awards JSON data: I imported this data into VS Code. This data seems reasonably fine, with no missing values in each attribute upon initial inspection.</p>

Date	Task	Procedure																																																																																																															
30/5	Cleaning CSV dataset in Excel	<p>Top 10K spotify artists csv data:</p> <ul style="list-style-type: none">- Redoing the index column to make it a unique identifier column for each artist. As seen below, it starts at 0 again so I will replace all the numbers with their position at the top10K + the first letter of their name (unique key). <table><tr><td>3164</td><td>3162</td><td>3162</td><td>GG</td><td>mixed</td><td>5</td><td>group</td></tr><tr><td>3165</td><td>3163</td><td>3163</td><td>Ilta</td><td>female</td><td>0</td><td>person</td></tr><tr><td>3166</td><td>3164</td><td>3164</td><td>Arvey</td><td></td><td>0</td><td></td></tr><tr><td>3167</td><td>3165</td><td>3165</td><td>Alli Neumann</td><td></td><td>0</td><td></td></tr><tr><td>3168</td><td>3166</td><td>0</td><td>Arvey</td><td></td><td>0</td><td></td></tr><tr><td>3169</td><td>3167</td><td>1</td><td>Alli Neumann</td><td></td><td>0</td><td></td></tr><tr><td>3170</td><td>3168</td><td>2</td><td>Saba</td><td>male</td><td>0</td><td>person</td></tr><tr><td>3171</td><td>3169</td><td>3</td><td>Pretenders</td><td>mixed</td><td>42</td><td>group</td></tr><tr><td>3172</td><td>3170</td><td>4</td><td>Celo & Abdi</td><td>mixed</td><td>0</td><td>group</td></tr></table> <table><thead><tr><th></th><th>index</th><th>key</th><th>artist</th></tr></thead><tbody><tr><td>0</td><td>0</td><td>0D</td><td>Drake</td></tr><tr><td>1</td><td>1</td><td>1P</td><td>Post Malone</td></tr><tr><td>2</td><td>2</td><td>2E</td><td>Ed Sheeran</td></tr><tr><td>3</td><td>3</td><td>3J</td><td>J Balvin</td></tr><tr><td>4</td><td>4</td><td>4B</td><td>Bad Bunny</td></tr><tr><td>5</td><td>5</td><td>5J</td><td>Justin Bieber</td></tr><tr><td>6</td><td>6</td><td>6O</td><td>Ozuna</td></tr><tr><td>7</td><td>7</td><td>7A</td><td>Ariana Grande</td></tr><tr><td>8</td><td>8</td><td>8K</td><td>Khalid</td></tr><tr><td>9</td><td>9</td><td>9D</td><td>Daddy Yankee</td></tr><tr><td>10</td><td>10</td><td>10T</td><td>Travis Scott</td></tr></tbody></table> <p>(seen to the left) Creating a unique identifier column ('key').</p> <p>Removing the index column also as it is no longer needed. Also naming the first column as 'position' on the Top 10K charts.</p>	3164	3162	3162	GG	mixed	5	group	3165	3163	3163	Ilta	female	0	person	3166	3164	3164	Arvey		0		3167	3165	3165	Alli Neumann		0		3168	3166	0	Arvey		0		3169	3167	1	Alli Neumann		0		3170	3168	2	Saba	male	0	person	3171	3169	3	Pretenders	mixed	42	group	3172	3170	4	Celo & Abdi	mixed	0	group		index	key	artist	0	0	0D	Drake	1	1	1P	Post Malone	2	2	2E	Ed Sheeran	3	3	3J	J Balvin	4	4	4B	Bad Bunny	5	5	5J	Justin Bieber	6	6	6O	Ozuna	7	7	7A	Ariana Grande	8	8	8K	Khalid	9	9	9D	Daddy Yankee	10	10	10T	Travis Scott
3164	3162	3162	GG	mixed	5	group																																																																																																											
3165	3163	3163	Ilta	female	0	person																																																																																																											
3166	3164	3164	Arvey		0																																																																																																												
3167	3165	3165	Alli Neumann		0																																																																																																												
3168	3166	0	Arvey		0																																																																																																												
3169	3167	1	Alli Neumann		0																																																																																																												
3170	3168	2	Saba	male	0	person																																																																																																											
3171	3169	3	Pretenders	mixed	42	group																																																																																																											
3172	3170	4	Celo & Abdi	mixed	0	group																																																																																																											
	index	key	artist																																																																																																														
0	0	0D	Drake																																																																																																														
1	1	1P	Post Malone																																																																																																														
2	2	2E	Ed Sheeran																																																																																																														
3	3	3J	J Balvin																																																																																																														
4	4	4B	Bad Bunny																																																																																																														
5	5	5J	Justin Bieber																																																																																																														
6	6	6O	Ozuna																																																																																																														
7	7	7A	Ariana Grande																																																																																																														
8	8	8K	Khalid																																																																																																														
9	9	9D	Daddy Yankee																																																																																																														
10	10	10T	Travis Scott																																																																																																														

Date	Task	Procedure
1/6	Cleaning dataset in Excel	<div>   </div> <p>Making position (in the charts) start from 1 and changing reference for key column. More logical this way.</p> <p>Age is an important column for answers to one of my questions, so I need to make sure that the ages of 'groups' is blank as it doesn't apply to them (because there are mixed ages in groups). Before I can do this however, I need to fix the blank values in the 'type' column as age is dependant on whether it is a group or a person. I changed the blank values to 'n/a' for type. The first column is 'type' (fixed column), and second is type_0 (original).</p> <div>   </div>
2/6	Cleaning dataset in Excel	<p>Made the age of groups dependant on what 'type' the artist's column is, for example the age of a group now will be 'n/a' opposed to a random number. All other ages entered in the age column for individuals are assumed valid.</p> <div>   </div> <div>  </div> <p>As seen here, the age column is now n/a for all groups (because there are multiple ages so we don't have a concrete age)</p>

Date	Task	Procedure
2/6	Fixing the grammy JSON data to get ready for combining datasets	<p>The grammy data has a very clean format already however, it is in one giant array which needs to be removed as MongoDB cannot parse an array like that.</p> <p>In my text editor Atom, I opened up the grammyData.json file and I removed the two square brackets <code>[]</code> that make the array. I did this quickly through find and replace query.</p> <ol style="list-style-type: none"> 1) Find <code>\[</code> 2) Replace (blank)  <p>I then tried to import through MongoDB but it didn't work because it says it can't parse the comma. So I removed all commas through find and replace query, but only the ones after a curly bracket.</p> <ol style="list-style-type: none"> 1) Find <code>{},</code> 2) Replace <code>\$1</code>  <p>Now it is ready to be imported into MongoDB via mongoimport.</p>
3/6	Importing Grammy JSON Data into MongoDB	  <p>Using the command line to import the grammy JSON data into MongoDB.</p>

Date	Task	Procedure																																																																																																																																																																																																											
3/6	Exporting JSON data into CSV format.	<div><div>Last login: Thu Jun 3 17:13:30 on ttys000</div><div>The default interactive shell is now zsh. To update your account to use zsh, please run `chsh -s /bin/zsh`. For more details, please visit https://support.apple.com/kb/HT208050. MacBook-Air-2:~ sumedhamukr\$ mongoexport --db=grammy --collection=awards --type=csv --fields=category,annualGrammy,awardType,name,awardFor --out=/opt/homebrew/grammys.csv 2021-06-03T21:01:35.683+1200 connected to: mongod://localhost/ 2021-06-03T21:01:35.756+1200 exported 4305 records MacBook-Air-2:~ sumedhamukr\$</div></div> <p>In the command shell I used the mongoexport command as shown above: mongoexport --db=grammy --collection=awards --type=csv --fields=category,annualGrammy,awardType,name,awardFor --out=/opt/homebrew/grammys.csv</p> <p>This created a file called grammys.csv in my /opt/homebrew/ folder. I quickly opened it in excel just to check it out and it looks good:</p> <table><tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th></tr><tr><th></th><th>category</th><th>annualGrammy</th><th>awardType</th><th>name</th><th>awardFor</th><th></th></tr><tr><td>1</td><td>Record Of The Year</td><td></td><td>1 Individual</td><td>Nel Blu Dipinto Di Blu (Volare)</td><td>Nel Blu Dipinto Di Blu (Volare)</td><td></td></tr><tr><td>2</td><td>Album Of The Year</td><td></td><td>1 Work</td><td>Henry Mancini</td><td>The Music From Peter Gunn</td><td></td></tr><tr><td>3</td><td>Best Jazz Performance, Individual</td><td></td><td>1 Work</td><td>Ella Fitzgerald</td><td>Ella Fitzgerald Sings The Duke Ellington Song Book</td><td></td></tr><tr><td>4</td><td>Best Comedy Performance</td><td></td><td>1 Individual</td><td>The Chipmunk Song</td><td>The Chipmunk Song</td><td></td></tr><tr><td>5</td><td>Best Jazz Performance, Group</td><td></td><td>1 Individual</td><td>Basie</td><td>Basie</td><td></td></tr><tr><td>6</td><td>Best Country & Western Performance</td><td></td><td>1 Individual</td><td>Tom Dooley</td><td>Tom Dooley</td><td></td></tr><tr><td>7</td><td>Best Rhythm & Blues Performance</td><td></td><td>1 Individual</td><td>Tequila</td><td>Tequila</td><td></td></tr><tr><td>8</td><td>Best Arrangement</td><td></td><td>1 Work</td><td>Henry Mancini</td><td>The Music From Peter Gunn</td><td></td></tr><tr><td>9</td><td>Best Engineered Record (Classical)</td><td></td><td>1 Individual</td><td>Duets With A Spanish Guitar</td><td>Duets With A Spanish Guitar</td><td></td></tr><tr><td>10</td><td>Best Vocal Performance, Male</td><td></td><td>1 Individual</td><td>Catch A Falling Star</td><td>Catch A Falling Star</td><td></td></tr><tr><td>11</td><td>Best Engineered Record - Non-Classical</td><td></td><td>1 Individual</td><td>The Chipmunk Song</td><td>The Chipmunk Song</td><td></td></tr><tr><td>12</td><td>Best Original Cast Album (Broadway Or TV)</td><td></td><td>1 Individual</td><td>The Music Man</td><td>The Music Man</td><td></td></tr><tr><td>13</td><td>Best Album Cover</td><td></td><td>1 Work</td><td>Frank Sinatra</td><td>Only The Lonely</td><td></td></tr><tr><td>14</td><td>Best Vocal Performance, Female</td><td></td><td>1 Work</td><td>Ella Fitzgerald</td><td>Ella Fitzgerald Sings The Irving Berlin Song Book</td><td></td></tr><tr><td>15</td><td>Best Recording For Children</td><td></td><td>1 Individual</td><td>The Chipmunk Song</td><td>The Chipmunk Song</td><td></td></tr><tr><td>16</td><td>Best Classical Performance - Orchestra</td><td></td><td>1 Individual</td><td>Gaite Parisienne</td><td>Gaite Parisienne</td><td></td></tr><tr><td>17</td><td>Best Musical Composition First Recorded And Released In 1958 (Over 5 Minutes Duration)</td><td></td><td>1 Individual</td><td>Cross Country Suite</td><td>Cross Country Suite</td><td></td></tr><tr><td>18</td><td>Best Classical Performance - Instrumentalist (Other Than Concerto-Scale Accompaniment)</td><td></td><td>1 Individual</td><td>Segovia Golden Jubilee</td><td>Segovia Golden Jubilee</td><td></td></tr><tr><td>19</td><td>Best Classical Performance - Instrumentalist (With Concerto Scale Accompaniment)</td><td></td><td>1 Individual</td><td>Tchaikovsky: Piano Concerto No. 1 In B F</td><td>Tchaikovsky: Piano Concerto No. 1 In B Flat Minor, Op. 23</td><td></td></tr><tr><td>20</td><td>Best Performance, Documentary Or Spoken Word</td><td></td><td>1 Individual</td><td>The Best Of The Stan Freberg Shows</td><td>The Best Of The Stan Freberg Shows</td><td></td></tr><tr><td>21</td><td>Best Performance By A Vocal Group Or Chorus</td><td></td><td>1 Individual</td><td>That Old Black Magic</td><td>That Old Black Magic</td><td></td></tr><tr><td>22</td><td>Song Of The Year</td><td></td><td>1 Individual</td><td>Nel Blu Dipinto Di Blu (Volare)</td><td>Nel Blu Dipinto Di Blu (Volare)</td><td></td></tr><tr><td>23</td><td>Best Classical Performance - Chamber Music (Including Chamber Orchestra)</td><td></td><td>1 Individual</td><td>Beethoven: Quartet 130</td><td>Beethoven: Quartet 130</td><td></td></tr><tr><td>24</td><td>Best Classical Performance - Vocal Soloist (With Or Without Orchestra)</td><td></td><td>1 Individual</td><td>Operatic Recital</td><td>Operatic Recital</td><td></td></tr><tr><td>25</td><td>Record Of The Year</td><td></td><td>2 Individual</td><td>Mack The Knife</td><td>Mack The Knife</td><td></td></tr><tr><td>26</td><td>Best Classical Performance - Operatic Or Choral</td><td></td><td>1 Individual</td><td>Virtuoso</td><td>Virtuoso</td><td></td></tr><tr><td>27</td><td>Album Of The Year</td><td></td><td>2 Work</td><td>Frank Sinatra</td><td>Come Dance With Me</td><td></td></tr></table>		A	B	C	D	E	F		category	annualGrammy	awardType	name	awardFor		1	Record Of The Year		1 Individual	Nel Blu Dipinto Di Blu (Volare)	Nel Blu Dipinto Di Blu (Volare)		2	Album Of The Year		1 Work	Henry Mancini	The Music From Peter Gunn		3	Best Jazz Performance, Individual		1 Work	Ella Fitzgerald	Ella Fitzgerald Sings The Duke Ellington Song Book		4	Best Comedy Performance		1 Individual	The Chipmunk Song	The Chipmunk Song		5	Best Jazz Performance, Group		1 Individual	Basie	Basie		6	Best Country & Western Performance		1 Individual	Tom Dooley	Tom Dooley		7	Best Rhythm & Blues Performance		1 Individual	Tequila	Tequila		8	Best Arrangement		1 Work	Henry Mancini	The Music From Peter Gunn		9	Best Engineered Record (Classical)		1 Individual	Duets With A Spanish Guitar	Duets With A Spanish Guitar		10	Best Vocal Performance, Male		1 Individual	Catch A Falling Star	Catch A Falling Star		11	Best Engineered Record - Non-Classical		1 Individual	The Chipmunk Song	The Chipmunk Song		12	Best Original Cast Album (Broadway Or TV)		1 Individual	The Music Man	The Music Man		13	Best Album Cover		1 Work	Frank Sinatra	Only The Lonely		14	Best Vocal Performance, Female		1 Work	Ella Fitzgerald	Ella Fitzgerald Sings The Irving Berlin Song Book		15	Best Recording For Children		1 Individual	The Chipmunk Song	The Chipmunk Song		16	Best Classical Performance - Orchestra		1 Individual	Gaite Parisienne	Gaite Parisienne		17	Best Musical Composition First Recorded And Released In 1958 (Over 5 Minutes Duration)		1 Individual	Cross Country Suite	Cross Country Suite		18	Best Classical Performance - Instrumentalist (Other Than Concerto-Scale Accompaniment)		1 Individual	Segovia Golden Jubilee	Segovia Golden Jubilee		19	Best Classical Performance - Instrumentalist (With Concerto Scale Accompaniment)		1 Individual	Tchaikovsky: Piano Concerto No. 1 In B F	Tchaikovsky: Piano Concerto No. 1 In B Flat Minor, Op. 23		20	Best Performance, Documentary Or Spoken Word		1 Individual	The Best Of The Stan Freberg Shows	The Best Of The Stan Freberg Shows		21	Best Performance By A Vocal Group Or Chorus		1 Individual	That Old Black Magic	That Old Black Magic		22	Song Of The Year		1 Individual	Nel Blu Dipinto Di Blu (Volare)	Nel Blu Dipinto Di Blu (Volare)		23	Best Classical Performance - Chamber Music (Including Chamber Orchestra)		1 Individual	Beethoven: Quartet 130	Beethoven: Quartet 130		24	Best Classical Performance - Vocal Soloist (With Or Without Orchestra)		1 Individual	Operatic Recital	Operatic Recital		25	Record Of The Year		2 Individual	Mack The Knife	Mack The Knife		26	Best Classical Performance - Operatic Or Choral		1 Individual	Virtuoso	Virtuoso		27	Album Of The Year		2 Work	Frank Sinatra	Come Dance With Me	
	A	B	C	D	E	F																																																																																																																																																																																																							
	category	annualGrammy	awardType	name	awardFor																																																																																																																																																																																																								
1	Record Of The Year		1 Individual	Nel Blu Dipinto Di Blu (Volare)	Nel Blu Dipinto Di Blu (Volare)																																																																																																																																																																																																								
2	Album Of The Year		1 Work	Henry Mancini	The Music From Peter Gunn																																																																																																																																																																																																								
3	Best Jazz Performance, Individual		1 Work	Ella Fitzgerald	Ella Fitzgerald Sings The Duke Ellington Song Book																																																																																																																																																																																																								
4	Best Comedy Performance		1 Individual	The Chipmunk Song	The Chipmunk Song																																																																																																																																																																																																								
5	Best Jazz Performance, Group		1 Individual	Basie	Basie																																																																																																																																																																																																								
6	Best Country & Western Performance		1 Individual	Tom Dooley	Tom Dooley																																																																																																																																																																																																								
7	Best Rhythm & Blues Performance		1 Individual	Tequila	Tequila																																																																																																																																																																																																								
8	Best Arrangement		1 Work	Henry Mancini	The Music From Peter Gunn																																																																																																																																																																																																								
9	Best Engineered Record (Classical)		1 Individual	Duets With A Spanish Guitar	Duets With A Spanish Guitar																																																																																																																																																																																																								
10	Best Vocal Performance, Male		1 Individual	Catch A Falling Star	Catch A Falling Star																																																																																																																																																																																																								
11	Best Engineered Record - Non-Classical		1 Individual	The Chipmunk Song	The Chipmunk Song																																																																																																																																																																																																								
12	Best Original Cast Album (Broadway Or TV)		1 Individual	The Music Man	The Music Man																																																																																																																																																																																																								
13	Best Album Cover		1 Work	Frank Sinatra	Only The Lonely																																																																																																																																																																																																								
14	Best Vocal Performance, Female		1 Work	Ella Fitzgerald	Ella Fitzgerald Sings The Irving Berlin Song Book																																																																																																																																																																																																								
15	Best Recording For Children		1 Individual	The Chipmunk Song	The Chipmunk Song																																																																																																																																																																																																								
16	Best Classical Performance - Orchestra		1 Individual	Gaite Parisienne	Gaite Parisienne																																																																																																																																																																																																								
17	Best Musical Composition First Recorded And Released In 1958 (Over 5 Minutes Duration)		1 Individual	Cross Country Suite	Cross Country Suite																																																																																																																																																																																																								
18	Best Classical Performance - Instrumentalist (Other Than Concerto-Scale Accompaniment)		1 Individual	Segovia Golden Jubilee	Segovia Golden Jubilee																																																																																																																																																																																																								
19	Best Classical Performance - Instrumentalist (With Concerto Scale Accompaniment)		1 Individual	Tchaikovsky: Piano Concerto No. 1 In B F	Tchaikovsky: Piano Concerto No. 1 In B Flat Minor, Op. 23																																																																																																																																																																																																								
20	Best Performance, Documentary Or Spoken Word		1 Individual	The Best Of The Stan Freberg Shows	The Best Of The Stan Freberg Shows																																																																																																																																																																																																								
21	Best Performance By A Vocal Group Or Chorus		1 Individual	That Old Black Magic	That Old Black Magic																																																																																																																																																																																																								
22	Song Of The Year		1 Individual	Nel Blu Dipinto Di Blu (Volare)	Nel Blu Dipinto Di Blu (Volare)																																																																																																																																																																																																								
23	Best Classical Performance - Chamber Music (Including Chamber Orchestra)		1 Individual	Beethoven: Quartet 130	Beethoven: Quartet 130																																																																																																																																																																																																								
24	Best Classical Performance - Vocal Soloist (With Or Without Orchestra)		1 Individual	Operatic Recital	Operatic Recital																																																																																																																																																																																																								
25	Record Of The Year		2 Individual	Mack The Knife	Mack The Knife																																																																																																																																																																																																								
26	Best Classical Performance - Operatic Or Choral		1 Individual	Virtuoso	Virtuoso																																																																																																																																																																																																								
27	Album Of The Year		2 Work	Frank Sinatra	Come Dance With Me																																																																																																																																																																																																								
4/6	Combining the datasets through Excel	<p>I have decided to combine the datasets through the excel sheet for the grammy data (seen previously). I have copy pasted the Top 10K sheet into a separate sheet in the same document.</p> <p>Note: I have only taken a subset of the Top 10K data (however I am using all the grammy data) because some of it is irrelevant to the questions I am asking. I am taking columns position, key, artist, gender, age and type.</p> <div><table><tr><td>43</td><td>42</td><td>42S</td><td>Sia</td></tr><tr><td>44</td><td>43</td><td>43S</td><td>Sech</td></tr><tr><td>45</td><td>44</td><td>44D</td><td>David Guetta</td></tr></table><div>grammys top10k +</div><div>Ready</div></div> <p>Preparing for VLOOKUP - by putting putting the ‘name’ column as the first column in the grammy sheet, and the ‘artist’ column as the first column in the top10k sheet. This is because they are the two I am going to link through VLOOKUP. (seen below)</p>	43	42	42S	Sia	44	43	43S	Sech	45	44	44D	David Guetta																																																																																																																																																																																															
43	42	42S	Sia																																																																																																																																																																																																										
44	43	43S	Sech																																																																																																																																																																																																										
45	44	44D	David Guetta																																																																																																																																																																																																										

Date

Task

Procedure

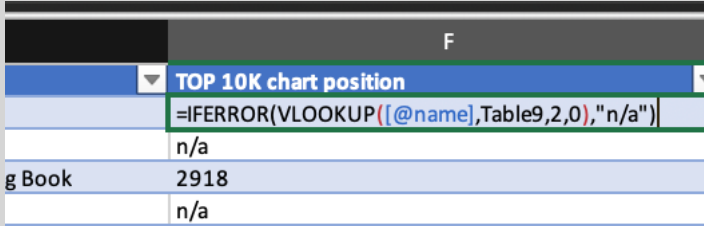
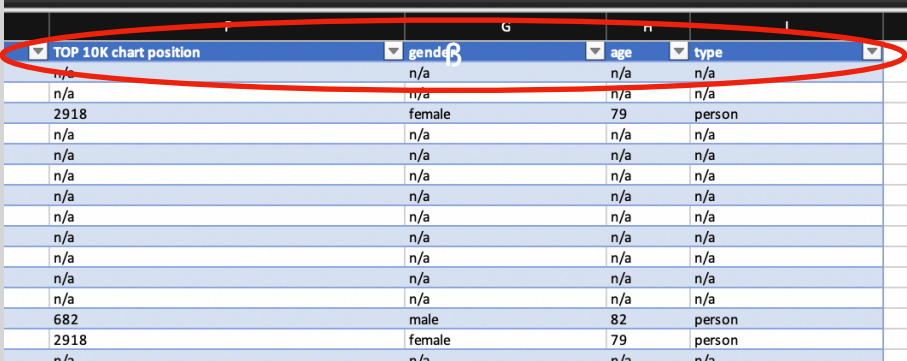
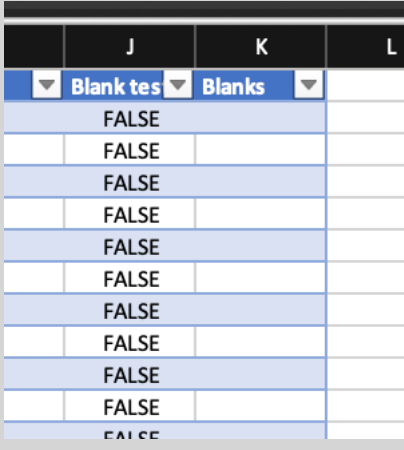
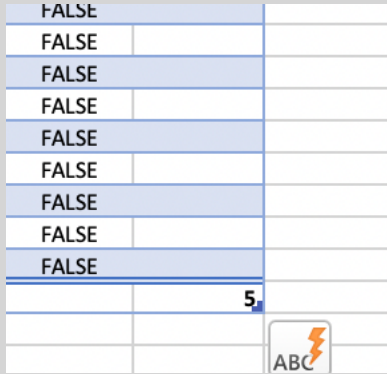
4/6

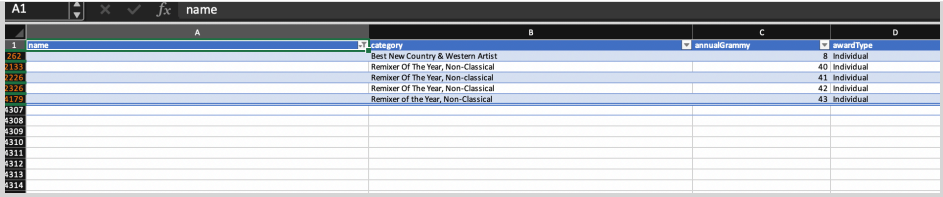
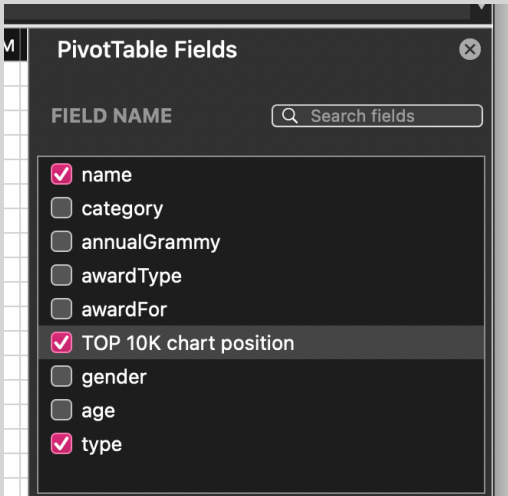
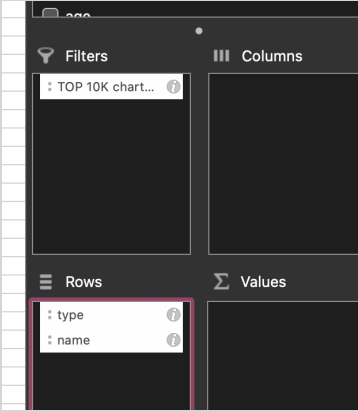
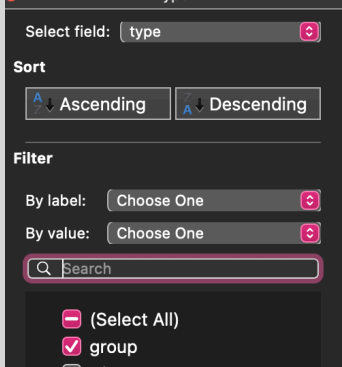
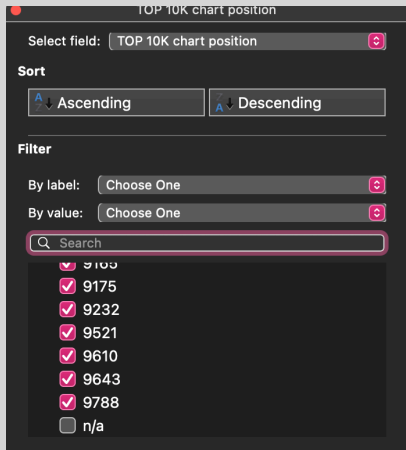
grammys sheet:

A		B		
name	category	annualGrammy		
Nel Blu Dipinto Di Blu (Volare)	Record Of The Year			
Henry Mancini	Album Of The Year			
Ella Fitzgerald	Best Jazz Performance, Individual			
The Chipmunk Song	Best Comedy Performance			
Basie	Best Jazz Performance, Group			
Tom Dooley	Best Country & Western Performance			
Tequila	Best Rhythm & Blues Performance			
Henry Mancini	Best Arrangement			
Duets With A Spanish Guitar	Best Engineered Record (Classical)			
Catch A Falling Star	Best Vocal Performance, Male			
The Chipmunk Song	Best Engineered Record - Non-Classical			
The Music Man	Best Original Cast Album (Broadway Or TV)			
Frank Sinatra	Best Album Cover			
Ella Fitzgerald	Best Vocal Performance, Female			
The Chipmunk Song	Best Recording For Children			
Gaite Parisienne	Best Classical Performance - Orchestra			
Cross Country Suite	Best Musical Composition First Recorded And Released In 1958 (Over 5 Minutes Duration)			
Segovia Golden Jubilee	Best Classical Performance - Instrumentalist (Other Than Concerto-Scale Accompaniment)			
Tchaikovsky: Piano Concerto No. 1 In B Flat Minor, Op. 23	Best Classical Performance - Instrumentalist (With Concerto Scale Accompaniment)			
The Best Of The Stan Freberg Shows	Best Performance, Documentary Or Spoken Word			
That Old Black Magic	Best Performance By A Vocal Group Or Chorus			
Nel Blu Dipinto Di Blu (Volare)	Song Of The Year			
Beethoven: Quartet 130	Best Classical Performance - Chamber Music (Including Chamber Orchestra)			
Operatic Recital	Best Classical Performance - Vocal Soloist (With Or Without Orchestra)			
Mack The Knife	Record Of The Year			
Virtuoso	Best Classical Performance - Operatic Or Choral			
Frank Sinatra	Album Of The Year			
The Battle Of New Orleans	Song Of The Year			
Ella Fitzgerald	Best Vocal Performance, Female			
Duke Ellington	Best Performance By A Dance Band			
Frank Sinatra	Best Vocal Performance, Male			
Battle Hymn Of The Republic	Best Performance By A Vocal Group Or Chorus			
Debussy: Images For Orchestra	Best Classical Performance - Orchestra			
Beethoven: Sonatas No. 21 In C (Waldstein) And No. 18 In E F	Best Classical Performance - Chamber Music (Including Chamber Orchestra)			
Mozart: The Marriage Of Figaro	Best Classical Performance - Opera Cast Or Choral			
Rachmaninoff: Piano Concerto No. 3	Best Classical Performance - Concerto Or Instrumental Soloist (With Full Orchestral Accompaniment)			
Bjoerling In Opera	Best Classical Performance - Vocal Soloist (With Or Without Orchestra)			
I Dig Chicks	Best Jazz Performance - Group			
Beethoven: Sonatas No. 21 In C (Waldstein) And No. 18 In E F	Best Classical Performance - Concerto Or Instrumental Soloist (Other Than Full Orchestral Accompaniment)			
Duke Ellington	Best Musical Composition First Recorded And Released In 1959 (More Than 5 Minutes Duration)			
Duke Ellington	Best Sound Track Album - Background Score From A Motion Picture Or Television			
Gypsy	Best Broadway Show Album			
Porgy And Bess	Best Sound Track Album, Original Cast - Motion Picture Or Television			
Inside Shelley Berman	Best Comedy Performance - Spoken Word			
The Battle Of Kookamonga	Best Comedy Performance - Musical			
A Lincoln Portrait	Best Performance - Documentary Or Spoken Word (Other Than Comedy)			
Midnight Flyer	Best Performance By A "Top 40" Artist			

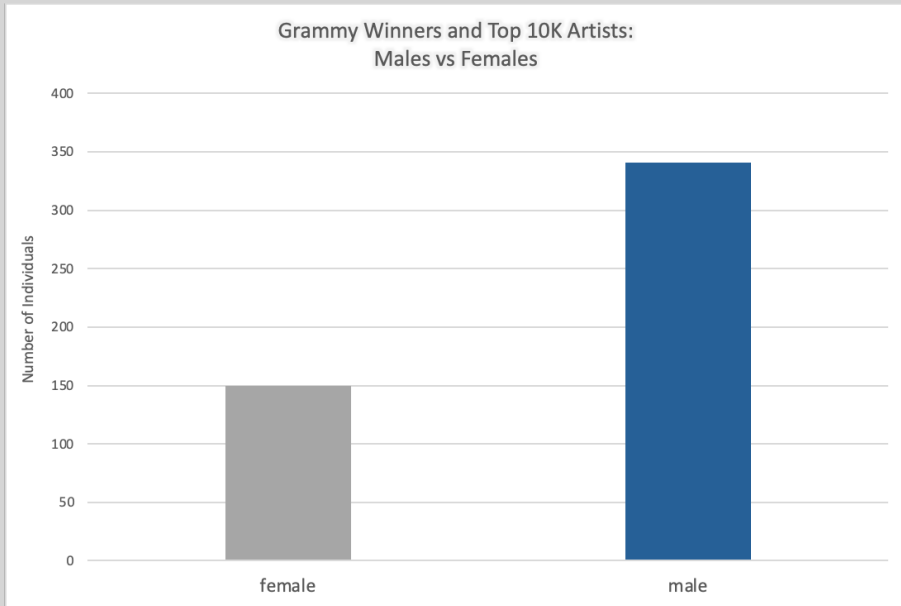
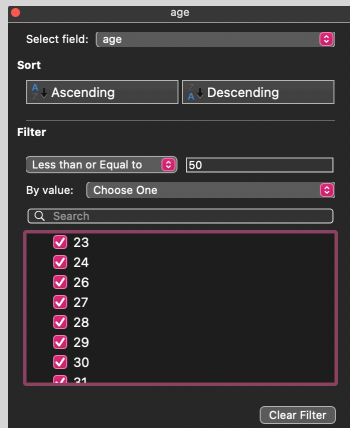
top10k sheet:

	A	B	C	D	
1	artist	position	key	gender	age
2	Drake	1	1D	male	
3	Post Malone	2	2P	male	
4	Ed Sheeran	3	3E	male	
5	J Balvin	4	4J	male	
6	Bad Bunny	5	5B	male	
7	Justin Bieber	6	6J	male	
8	Ozuna	7	7O	male	
9	Ariana Grande	8	8A	female	
10	Khalid	9	9K	male	
11	Daddy Yankee	10	10D	male	
12	Travis Scott	11	11T	male	
13	The Weeknd	12	12T	male	
14	XXXTENTACION	13	13X	male	
15	Billie Eilish	14	14B	female	
16	The Chainsmokers	15	15T	mixed	
17	Anuel Aa	16	16A	male	
18	Nicky Jam	17	17N	male	
19	Maluma	18	18M	male	
20	Camila Cabello	19	19C	female	
21	Dua Lipa	20	20D	female	
22	Juice WRLD	21	21J	male	
23	Selena Gomez	22	22S	female	
24	Halset	23	23H	female	
25	Cardi B	24	24C	female	
26	Calvin Harris	25	25C	male	
27	Nicki Minaj	26	26N	female	
28	Shawn Mendes	27	27S	male	
29	Imagine Dragons	28	28I	mixed	
30	Kendrick Lamar	29	29K	male	
31	Sam Smith	30	30S	other	
32	Future	31	31F	male	
33	Farruko	32	32F	male	
34	Maroon 5	33	33M	mixed	
35	Rihanna	34	34R	female	
36	Lil Uzi Vert	35	35L	male	

Date	Task	Procedure
5/6	Combining both datasets in Excel using VLOOKUP	<p>I used a VLOOKUP formula to connect the name columns to get the position of that artists on the spotify charts. (seen below). If they weren't found in the top10K, they were labelled as 'n/a'.</p>  <p>This was repeated for columns gender, age and type in the top 10k sheet.</p>  <p>^ The four new columns added to the grammy data. Therefore the data is now combined.</p>
6/6	Re-auditing data and checking for anything else that needs cleaning.	<p>To ensure my data is clean, I used the ISBLANK() function on the 'name' column in the combined data store. ('blank test' and 'blanks') where the blanks column displays a 1 where is blank is true. I then auto summed that column and there were actually 5 rows that were blank in name.</p>   <p>I also did this for all the columns I hadn't checked in the grammy data beforehand, and the only columns with blanks seemed to be name and awardFor, with 5 and 6 blanks respectively.</p> <ul style="list-style-type: none"> - To fix this issue, I filtered out the blanks in each column and deleted those rows. They are not useful when combined with the spotify data if they don't have names!

Date	Task	Procedure
6/6		 <p>^ filtered and deleted. Did not delete the extra 1 that was blank for awardFor, as it still had a name which is relevant.</p>
7/6	Wrangling data to achieve question of interest 1	<p>Question 1: How many groups have won at least one grammy and also are in the top 10K spotify artists?</p>   <p>The only way to achieve a unique count of the artists who have won Grammys (as their names can be listed more than once due to winning multiple grammys) while also showing their chart position was to put the two columns in the 'rows' section of the pivot table.</p> <p>I aggregated by group, then by name and filtered by chart position. Upon counting, I identified 53 unique groups that have won grammy(s) and also are in the top 10K charts.</p> <p>Note: aggregated group by filtering type by only group, aggregated Top 10K chart position by filtering out n/a.</p>  

Date	Task	Procedure																																																			
7/6		<p>I aggregated by group, then by name and chart position. Upon counting, I identified 53 unique groups that have won grammy(s) and also are in the top 10K charts. I put the results of all groups who have won grammys against all individuals who have won grammys (who are in the top 10K spotify artists list) in a pie chart.</p> <table><tr><th></th><th>A</th><th>B</th></tr><tr><td>1</td><td>TOP 10K chart position</td><td>(Multiple Items) -T</td></tr><tr><td>2</td><td></td><td></td></tr><tr><td>3</td><td>Row Labels</td><td>-T</td></tr><tr><td>4</td><td>group</td><td></td></tr><tr><td>5</td><td>A Great Big World</td><td></td></tr><tr><td>6</td><td>Aerosmith</td><td></td></tr><tr><td>7</td><td>Alabama Shakes</td><td></td></tr><tr><td>8</td><td>Arcade Fire</td><td></td></tr><tr><td>9</td><td>Blue</td><td></td></tr><tr><td>10</td><td>Bon Iver</td><td></td></tr><tr><td>11</td><td>Bon Jovi</td><td></td></tr><tr><td>12</td><td>Brothers</td><td></td></tr><tr><td>13</td><td>Busted</td><td></td></tr><tr><td>14</td><td>Calle 13</td><td></td></tr></table> <div><p>Grammy winners who are a Top 10K Spotify Artist</p><table><thead><tr><th>Category</th><th>Percentage</th></tr></thead><tbody><tr><td>group</td><td>73%</td></tr><tr><td>person</td><td>27%</td></tr></tbody></table></div>		A	B	1	TOP 10K chart position	(Multiple Items) -T	2			3	Row Labels	-T	4	group		5	A Great Big World		6	Aerosmith		7	Alabama Shakes		8	Arcade Fire		9	Blue		10	Bon Iver		11	Bon Jovi		12	Brothers		13	Busted		14	Calle 13		Category	Percentage	group	73%	person	27%
	A	B																																																			
1	TOP 10K chart position	(Multiple Items) -T																																																			
2																																																					
3	Row Labels	-T																																																			
4	group																																																				
5	A Great Big World																																																				
6	Aerosmith																																																				
7	Alabama Shakes																																																				
8	Arcade Fire																																																				
9	Blue																																																				
10	Bon Iver																																																				
11	Bon Jovi																																																				
12	Brothers																																																				
13	Busted																																																				
14	Calle 13																																																				
Category	Percentage																																																				
group	73%																																																				
person	27%																																																				
8/6	Wrangling to achieve question of interest 2	<p>Question 2: Of those who are in the Top 10K artists on Spotify, have females won more Grammys than males?</p> <p>Note: upon starting the pivot table, it showed that there was a gender named '0' so I investigated this further by filtering the gender column in the combined dataset to see how many rows had this value. It was only one, however, so I just manually changed this value to 'n/a'.</p> <div><table><tr><th colspan="2">Filters</th><th colspan="2">Columns</th></tr><tr><td colspan="2">TOP 10K chart...</td><td colspan="2"></td></tr></table><table><tr><th colspan="2">Rows</th><th colspan="2">Values</th></tr><tr><td colspan="2">gender</td><td colspan="2">Count of name</td></tr></table><table><tr><th>A</th><th>B</th></tr><tr><td>TOP 10K chart position</td><td>(Multiple Items) -T</td></tr><tr><td>Row Labels</td><td>Count of name</td></tr><tr><td>female</td><td>150</td></tr><tr><td>male</td><td>341</td></tr><tr><td>Grand Total</td><td>491</td></tr></table></div> <p>The answer appears to be no; more males have won Grammys and also been in the top 10K.</p>	Filters		Columns		TOP 10K chart...				Rows		Values		gender		Count of name		A	B	TOP 10K chart position	(Multiple Items) -T	Row Labels	Count of name	female	150	male	341	Grand Total	491																							
Filters		Columns																																																			
TOP 10K chart...																																																					
Rows		Values																																																			
gender		Count of name																																																			
A	B																																																				
TOP 10K chart position	(Multiple Items) -T																																																				
Row Labels	Count of name																																																				
female	150																																																				
male	341																																																				
Grand Total	491																																																				

Date	Task	Procedure						
8/6		<p>This is the chart I have come up with:</p> <div><p>Grammy Winners and Top 10K Artists: Males vs Females</p><table><thead><tr><th>Gender</th><th>Number of Individuals</th></tr></thead><tbody><tr><td>female</td><td>~150</td></tr><tr><td>male</td><td>~340</td></tr></tbody></table></div>	Gender	Number of Individuals	female	~150	male	~340
Gender	Number of Individuals							
female	~150							
male	~340							
8/6	Wrangling data to achieve question 3 of interest	<p>Question 3: Which Grammy Award category has the most winners equal to or under the age of 50?</p> <p>Note: Encountered same problem as question 2, found four rows with age set to '0'. Manually changed them to n/a. (if it was more than four would have used a formula!)</p> <p>1. Firstly sorted by category first, then age (both in rows section) and then put age in values.</p> <div></div> <p>As seen in the right screenshot, I filtered age so that it was less than or equal to 50. This resulted in an individual count of the number of ages under 50 for each category and their total. When sorted ascendingly, this gave 'Best Rap Album' as the answer, with 15 awards going to people under the age or equal to 50.</p>						

Date	Task	Procedure																																																																		
8/6		<p>As seen to the right, it shows the different age winner groups for each Grammy category, Best Rap Album has the most!</p> <table><tr><th>Row Labels</th><th>Count of age</th></tr><tr><td>Best Rap Album</td><td>15</td></tr><tr><td>47</td><td>6</td></tr><tr><td>43</td><td>4</td></tr><tr><td>37</td><td>1</td></tr><tr><td>33</td><td>3</td></tr><tr><td>27</td><td>1</td></tr><tr><td>Best New Artist</td><td>13</td></tr><tr><td>50</td><td>1</td></tr><tr><td>41</td><td>2</td></tr><tr><td>39</td><td>2</td></tr><tr><td>37</td><td>1</td></tr><tr><td>35</td><td>1</td></tr><tr><td>32</td><td>1</td></tr><tr><td>28</td><td>1</td></tr><tr><td>27</td><td>2</td></tr><tr><td>26</td><td>1</td></tr><tr><td>24</td><td>1</td></tr><tr><td>Best Pop Vocal Album</td><td>13</td></tr><tr><td>42</td><td>1</td></tr><tr><td>41</td><td>1</td></tr><tr><td>39</td><td>1</td></tr><tr><td>38</td><td>2</td></tr><tr><td>34</td><td>2</td></tr><tr><td>32</td><td>2</td></tr><tr><td>30</td><td>1</td></tr><tr><td>29</td><td>1</td></tr><tr><td>28</td><td>1</td></tr><tr><td>27</td><td>1</td></tr><tr><td>Best Rap Song</td><td>11</td></tr><tr><td>47</td><td>3</td></tr><tr><td>43</td><td>4</td></tr><tr><td>39</td><td>1</td></tr></table>	Row Labels	Count of age	Best Rap Album	15	47	6	43	4	37	1	33	3	27	1	Best New Artist	13	50	1	41	2	39	2	37	1	35	1	32	1	28	1	27	2	26	1	24	1	Best Pop Vocal Album	13	42	1	41	1	39	1	38	2	34	2	32	2	30	1	29	1	28	1	27	1	Best Rap Song	11	47	3	43	4	39	1
Row Labels	Count of age																																																																			
Best Rap Album	15																																																																			
47	6																																																																			
43	4																																																																			
37	1																																																																			
33	3																																																																			
27	1																																																																			
Best New Artist	13																																																																			
50	1																																																																			
41	2																																																																			
39	2																																																																			
37	1																																																																			
35	1																																																																			
32	1																																																																			
28	1																																																																			
27	2																																																																			
26	1																																																																			
24	1																																																																			
Best Pop Vocal Album	13																																																																			
42	1																																																																			
41	1																																																																			
39	1																																																																			
38	2																																																																			
34	2																																																																			
32	2																																																																			
30	1																																																																			
29	1																																																																			
28	1																																																																			
27	1																																																																			
Best Rap Song	11																																																																			
47	3																																																																			
43	4																																																																			
39	1																																																																			

Link to final combined datastore: <https://drive.google.com/file/d/1DpWHIEligbYqRsAfrNRQ76-Bd6AdjPm8/view?usp=sharing>

iii) Answers

Question 1:

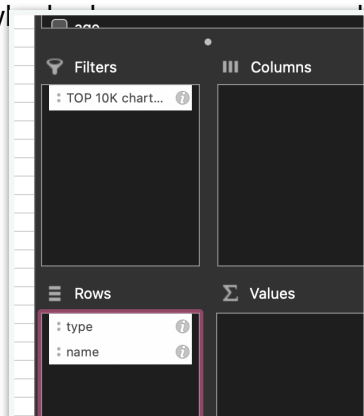
How many groups have won at least one Grammy and also are in the Top 10K Spotify artists?

Conclusion:

53 unique groups have one at least one Grammy and also are in the Top 10K artists on Spotify, although in general out of all (non-unique) groups/individuals who have won Grammy awards, 27% of them are groups.

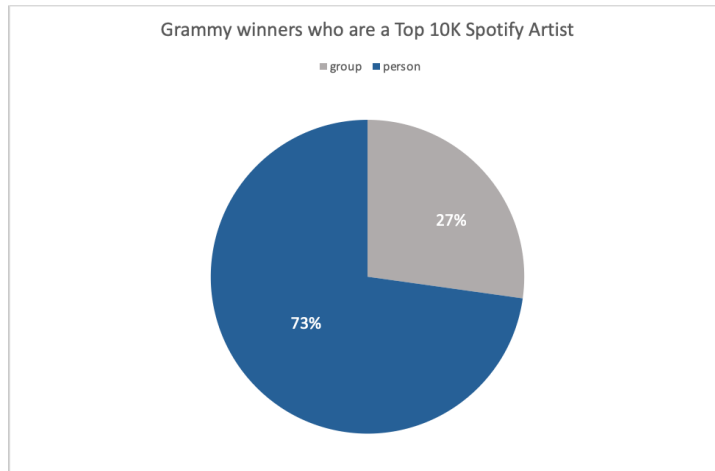
Steps taken:

- 1) Starting with Microsoft Excel, I created a PivotTable on a separate sheet, aggregating first by type and and name in the 'rows' section.
- 2) I then filtered by Top 10K artist position (because only those who have a position next to their name are in the top 10K), while also filtering out n/a values. This gave the list of unique groups with Grammy awards, which were then counted 53.



A	B
TOP 10K chart position	(Multiple Items)
Row Labels	
group	
A Great Big World	
Aerosmith	
Alabama Shakes	
Arcade Fire	
Blue	
Bon Iver	
Bon Jovi	
Brothers	
Busted	
Call Me	

3) To get a graph of all the 'types' of artists (groups/persons), I created a pivot chart of Group vs. Person and added % labels to the pie chart.



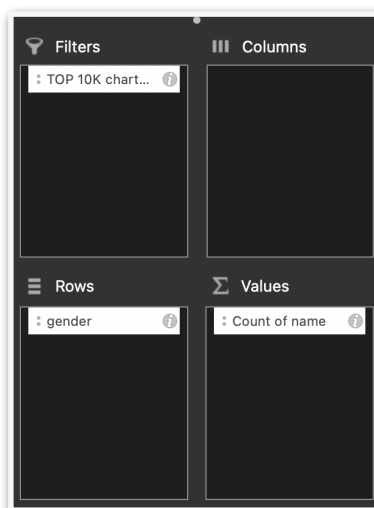
Question 2:

Of those who are in the Top 10K artists on Spotify, have females won more Grammys than males?

Conclusion: After analysis, the answer is no, as males have won more Grammys than females while also being in the Top 10K artists on Spotify.

Steps taken:

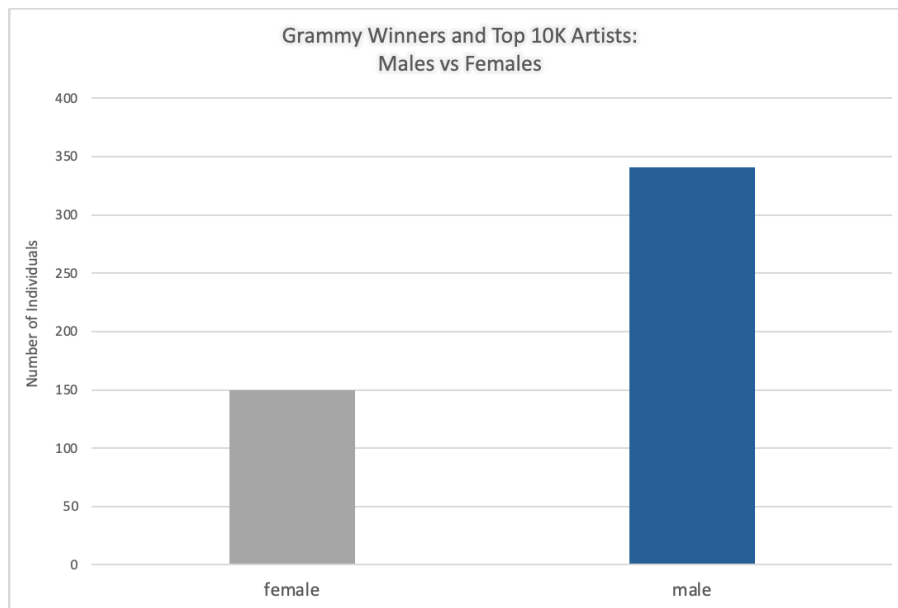
- 1) I opened a new PivotTable in Microsoft Excel based on my combined dataset.
- 2) First filter by Top 10K artist chart position
- 3) Place gender in the rows section, and values is the count of 'name' column (hence counting all the names in each gender)



A	B
TOP 10K chart position	(Multiple Items)
Row Labels	Count of name
female	150
male	341
Grand Total	491

4) It is shown above that females have a much lower count of names (who have won Grammys) than males in the Top 10K. Specifically, of males who are a Top 10K Spotify artist, they have won 191 more Grammys overall than females who are a Top 10K Spotify artist.

5) To create a visual representation of this, I created a bar chart through the 'PivotChart' function on Microsoft Excel, and added a vertical axis label.



Question 3:

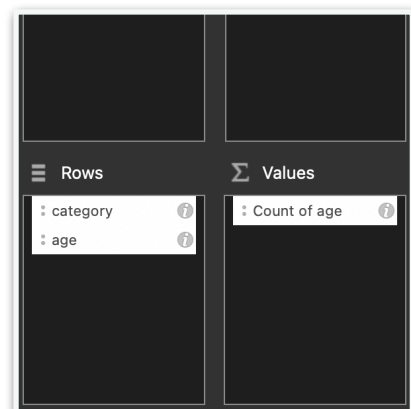
Which Grammy Award category has the most winners equal to or under the age of 50?

Conclusion: After analysis, the answer for which Grammy Award category has the most winners equal to or under the age of 50 is the category 'Best Rap Album', with 15 awards.

Steps taken:

- 1) I created another PivotChart in Microsoft Excel
- 2) I filtered category and age (in that order) by rows, and put age again in the values (so it will count the ages I will filter next)
- 3) To filter it for less than or equal to 50, I sorted the age in the filter menu 'By label' and set it to 'Less than or equal to' : 50. I then sorted again descendingly to get the highest number.

Row Labels	Count of age
Best Rap Album	15
47	6
43	4
37	1
33	3
27	1
Best New Artist	13
50	1
41	2
39	2
37	1
35	1
32	1
28	1
27	2
26	1
24	1
Best Pop Vocal Album	13
42	1
41	1
39	1
38	2
34	2
32	2
30	1
29	1
28	1
27	1
Best Rap Song	11
47	3
43	4
39	1



Question 4:

Of the Top 100 Artists in Spotify's Top 10K (who have won Grammy awards), is their number of Grammy awards higher depending on their popularity on Spotify?

Conclusion: As seen from the chart and analysis below, it is clear that just because an artist is more popular on Spotify does not mean that they get a higher amount of Grammy awards. However, it is evident that if an artist is in the Top 100 Artists on Spotify or similar, then having at least one Grammy award is not surprising.

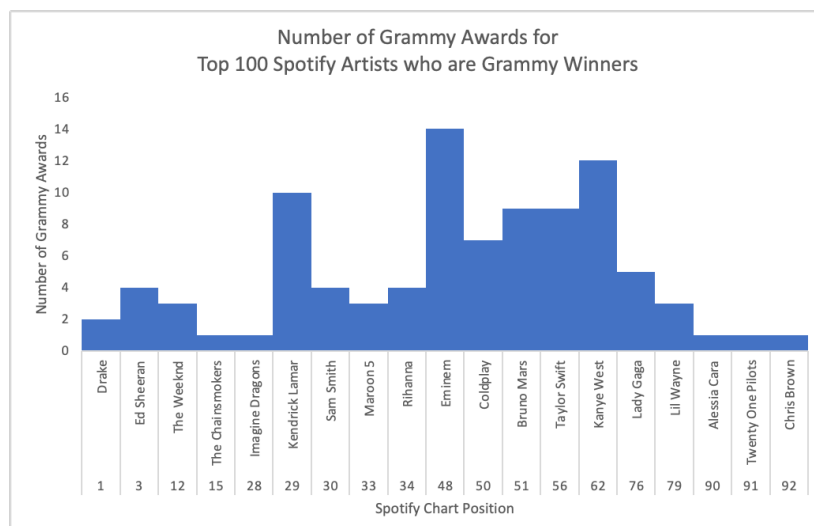
Steps taken:

- 1) Creating a PivotTable in Microsoft Excel, I put Top 10K chart position first in rows, and subsequently put 'name' also in rows.
- 2) After this, I put name in 'values' to count the number of names (therefore, awards)
- 3) I changed to the Top 10K chart position filter to select only chart positions ≥ 100 .
- 4) I sorted Top 10K chart position ascendingly (to see who was most popular). This created this visualisation:

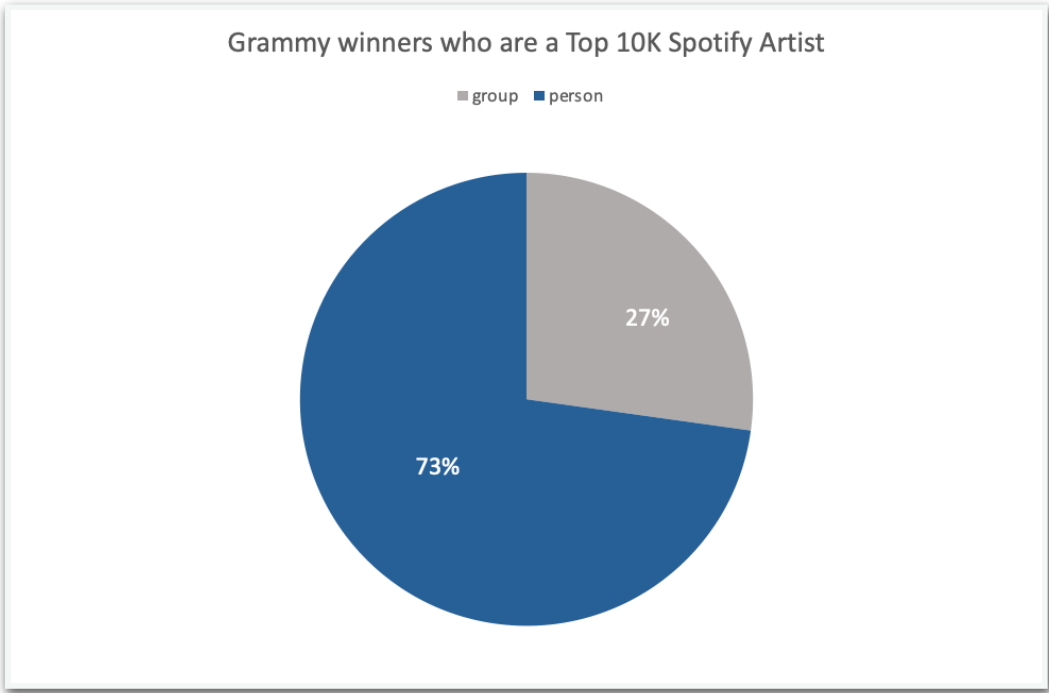
Row Labels	Count of name
1	2
Drake	2
3	4
Ed Sheeran	4
12	3
The Weeknd	3
15	1
The Chainsmokers	1
28	1
Imagine Dragons	1
29	10
Kendrick Lamar	10
30	4
Sam Smith	4
33	3
Maroon 5	3
34	4
Rihanna	4
48	14
Eminem	14
50	7

This shows (for example) Drake, who is #1 in the Top 100 (and top 10K) on Spotify but has only 2 Grammys (count of name).

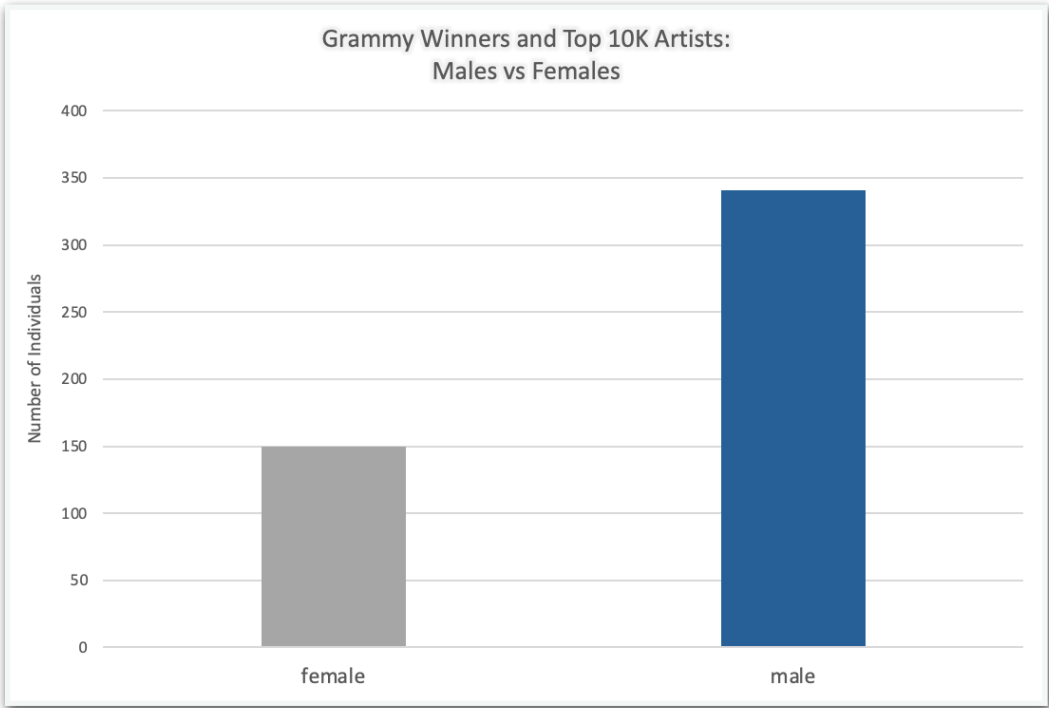
- 5) I created a PivotChart using the PivotTable and used the default chart, but I added some axis labels and changed the theme.



Appendices



Appendix 2



Appendix 3

