

Report for Databases and Advanced Data Techniques

Student: Cristina DeLisle

Student number: 190390475

1. Aspects regarding the dataset

1.1 Source of the dataset

The datasource for the dataset is the Keggles platform, originating in a data science project shared publicly with the scientist community. The link is the following: <https://www.kaggle.com/prajitr/eurovision-2021-song-contest-analysis/data> . For the present analysis, I have used the two sets present on this platform:

- Eurovision_juryvotes_2021.csv
- Eurovision_televotes_2021.csv

As the source of the data is public and there is no license constraint in using it, it is appropriate for the current project, from an ethical point of view. However, it proved to be too complex for the purposes of this project, so it was necessary to clean it in order to make it ready to be used in the application.

The tool used to do this was Google spreadsheets. I have uploaded the CSV files, one by one, in Drive and opened them with Google spreadsheets. Here I have done the following changes:

- Changing the name of the first column into 'Finalists'. This made more sense for what I planned to do further.
- Transposing the columns and the rows in such a way that the 26 Finalists are listed in the first column only, on each row instead of having separate columns for each. The countries that offered points (39 in total) represent the ones that participated in the competition, but did not qualify for the final. They are offering points to each of the Finalists: both coming from a professional jury and the general public.
- I have made a separate Column with Total Juryvotes, respectively Total Televotes in each of the tables and deleted all the other detailed entries with all the 39 participants. I kept only the sum of them. This decision was made considering that the current project does not need that level of detail. I am only interested in the Total points that were awarded in the Juryvotes and the Televotes during the competition.

1.2. Assessment of the dataset

In terms of quality of the dataset, we can notice that the data used comes from a reliable source and reflects the results that were given in the Eurovision 2021 song competition. It would have been better if there would have been a single CSV dataset that I could have used, but because the competition happened recently, the number of public sources that give using rights on the

dataset are limited. However, it is providing the relevant information for this project, which is to have an accurate overview of the results from the juryvotes and the televotes in the Eurovision 2021 competition.

In terms of the level of detail, the original CSV files were too detailed for the purposes of this project, so in the cleaning of the data it was necessary to sum values and delete columns.

The documentation aspect has as source the actual competition results, which are public and of notoriety. It was not necessary to make a bibliography for the data or anything in that matter.

In terms of interrelation, the aspects are presented in an objective way, connected directly to the official, public results from the competition. Both the juryvotes and the televotes are connected to each other by them being offered to the same country in part, culminating in the final results of the competition.

As usage, the data is widely used for research regarding the competition and is public.

The discoverability of the data is limited to the results of the competition and does not constitute a novelty element or some scientific discovery specific to it.

1.3 Interest in the dataset

The dataset is interesting from the perspective of knowing what results were obtained this year at the Eurovision 2021 competition. They reflect juryvotes and televotes, both representing Europe's view regarding the overall results: one being a qualitative score offered by a professional jury and another being a quantitative result, coming from the audience. The interest in the dataset is to point out the total scores obtained and the questions I am looking for to answer in this project are the following:

- Who were the finalists that qualified in the competition?
- What was the overall result for the juryvote, in a descending order?
- What was the overall result in the televote, in a descending order?
- What were the final results summing both the juryvotes and the televotes, in a descending order?

Answering these questions will point out the final results and also will be a great way to see if they are identical or if there are some differences in the way the jury and the public voted.

2. Modeling the data

2.1 E/R model of the data

The E/R diagram for the database is seen in Image 1.

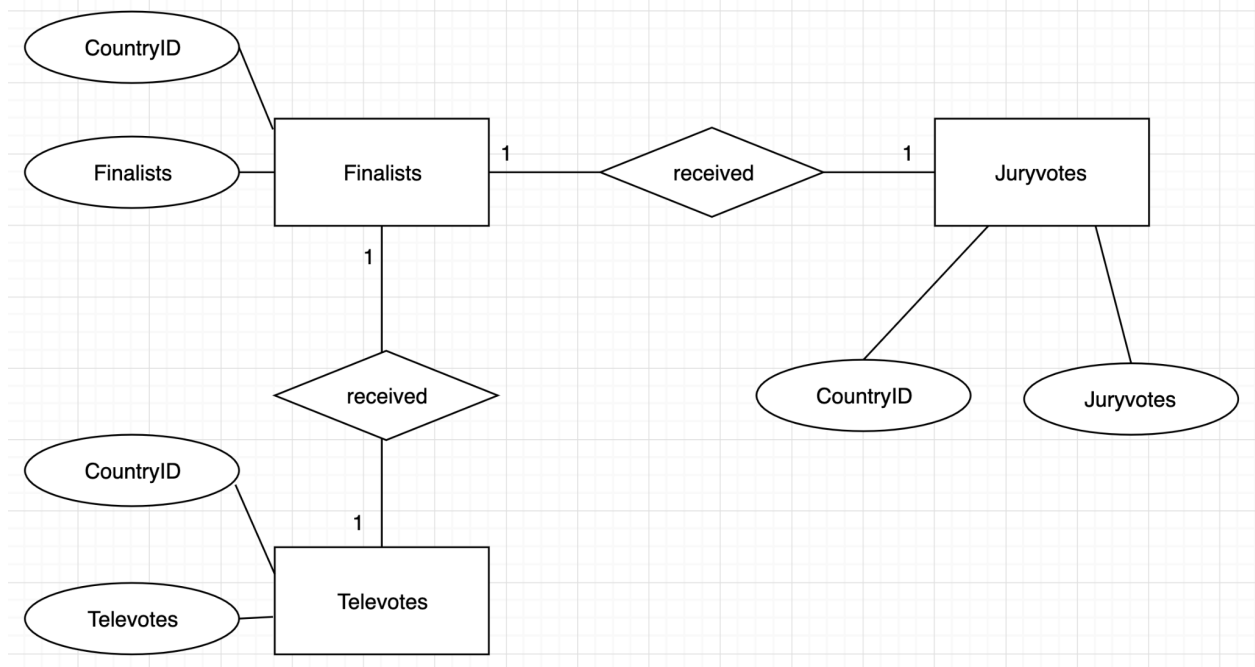


Image 1

The diagram represents the three tables which will constitute the database structure: finalists, juryvotes and televotes.

Compared to the dataset which was cleaned, the data reflects the dataset accurately, given the modifications that were done in the original CSV files that contained further details on the points that each country has awarded to the finalists. However, they are presented differently, in three tables, for normalization modeling purposes and adjustments necessary for the relational database. The database will be called resultsEurovision.

2.2. Aspects regarding the cardinality of the E/R diagram

We can notice in the Image 1 that the cardinality from Finalists and Juryvotes is 1 to 1. This is because every country from the rows of the Finalists table has received a separate entry in the rows of the Juryvotes table. Also, the cardinality from Finalists and Televotes is 1 to 1. This is because every country from the rows of the Finalists table has received a separate entry in the rows of the Televotes table.

However, in the dataset we can notice some cases where countries received the same point: e.g. the case of UK, Netherlands, Spain and Germany receiving 0 points in the Televotes. In this case, the relation is not modified, because every row from the tables has a 1 to 1 relationship: every country has received separate points, exactly one result in the juryvotes and one result in the televotes, which are relevant for calculating the final results.

2.3. Database tables and fields. Evaluation of the tables against the normal forms

According to the Image 1, the database has the following structure:

- Three tables: Finalists, Televotes and Juryvotes. The finalists table is connected to both the other tables, given that the Finalists received points from Televotes and Juryvotes.
- The Finalists table has a CountryID and a Finalists field. The CountryID is a PRIMARY KEY.
- The Televotes table has a CountryID and a Televotes field. The CountryID is a PRIMARY KEY.
- The Juryvotes table has a CountryID and a Juryvotes field. The CountryID is a PRIMARY KEY.

From the perspective of the normal forms, we can notice the following aspects:

- The tables are NF1, because each column is unique:
 - The Finalists table has two unique columns: a CountryID and a Finalists one.
 - The Juryvotes table has two unique columns: a CountryID and a Juryvotes one.
 - The Televotes table has two unique columns: a CountryID and a Televotes one.
- The tables are NF2, because they are NF1 and also all attributes within one entity depend solely on the unique identifier of that entity.
 - The Finalists table has two unique columns: a CountryID and a Finalists one, both depending on each other on the unique identifier of the CountryID. Every country is listed once and every identifier is corresponding uniquely to one country.
 - The Juryvotes table has two unique columns: a CountryID and a Juryvotes one, both depending on each other on the unique identifier of the CountryID. Every country is listed once and every identifier is corresponding uniquely to one country.
 - The Televotes table has two unique columns: a CountryID and a Televotes one, both depending on each other on the unique identifier of the CountryID. Every country is listed once and every identifier is corresponding uniquely to one country.
- The tables are NF3, because they are NF2 and no column entry is dependent on any other entry (value) other than the key for the respective table.
 - The CountryID represents the PRIMARY KEY in the table Finalists and no column entry from the Finalists column corresponds to another entry from the table apart from the CountryID
 - The CountryID represents the PRIMARY KEY in the table Juryvotes and no column entry from the Juryvotes column corresponds to another entry from the table apart from the CountryID
 - The CountryID represents the PRIMARY KEY in the table Televotes and no column entry from the Televotes column corresponds to another entry from the table apart from the CountryID
- The tables are not BCNF or 4NF or 5NF, given that the entire database resultsEurovision does not have the same unique key. Even if CountryID is the name of the PRIMARY

KEY in all three tables, these values, though coinciding, are specific to each table, given the definition of the PRIMARY KEY.

Overall, given that the tables are 3NF, the database results Eurovision is considered normalized. It's not necessary to go further with the normalization, given that the data is granular enough for the purposes of the project and a further normalization would not bring benefits.

3. Database creation

3.1 Database structure and CREATE commands

The database is structured into three tables:

- Finalists table, containing the CountryID column and the Finalists column, presented in alphabetical order. This can be seen in Image 2.
- Juryvotes table, containing the CountryID column and the Juryvotes column. This can be seen in Image 3.
- Televotes table, containing the CountryID column and the Televotes column. This can be seen in Image 4.

```
mysql> SELECT * from finalists;
+-----+-----+
| CountryID | Finalists |
+-----+-----+
| 1 | Albania |
| 2 | Azerbaijan |
| 3 | Belgium |
| 4 | Bulgaria |
| 5 | Cyprus |
| 6 | Finland |
| 7 | France |
| 8 | Germany |
| 9 | Greece |
| 10 | Iceland |
| 11 | Israel |
| 12 | Italy |
| 13 | Lithuania |
| 14 | Malta |
| 15 | Moldova |
| 16 | Netherlands |
| 17 | Norway |
| 18 | Portugal |
| 19 | Russia |
| 20 | San Marino |
| 21 | Serbia |
| 22 | Spain |
| 23 | Sweden |
| 24 | Switzerland |
| 25 | Ukraine |
| 26 | United Kingdom |
+-----+-----+
26 rows in set (0.00 sec)
```

Image 2

```
mysql> SELECT * FROM juryvotes;
```

CountryID	Juryvotes
1	22
2	32
3	71
4	140
5	50
6	83
7	248
8	3
9	91
10	198
11	73
12	206
13	55
14	208
15	53
16	11
17	15
18	126
19	104
20	37
21	20
22	6
23	46
24	267
25	97
26	0

```
26 rows in set (0.00 sec)
```

Image 3

```
mysql> SELECT * FROM televotes;
```

CountryID	Televotes
1	35
2	33
3	3
4	30
5	44
6	218
7	251
8	0
9	79
10	180
11	20
12	318
13	165
14	47
15	62
16	0
17	60
18	27
19	100
20	13
21	82
22	0
23	63
24	165
25	267
26	0

```
26 rows in set (0.00 sec)
```

Image 4

In terms of the CREATE commands, they are as follows:

- For creating the database resultsEurovision, as seen in Image 5:

```
CREATE DATABASE resultsEurovision;
```

- For creating the table finalists, as seen in Image 6:

```
CREATE TABLE finalists (  
  CountryID int NOT NULL AUTO_INCREMENT,  
  Finalists VARCHAR(255) NOT NULL,  
  PRIMARY KEY(CountryID)  
);
```

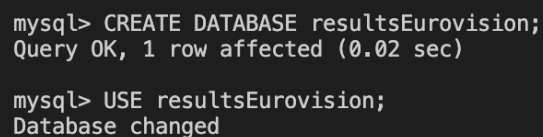
- For creating the table juryvotes, as seen in Image 7:

```
CREATE TABLE juryvotes (  
  CountryID int NOT NULL AUTO_INCREMENT,  
  Juryvotes DECIMAL (3, 0) NOT NULL,  
  PRIMARY KEY(CountryID)  
);
```

- For creating the table televotes, as seen in Image 8:

```
CREATE TABLE televotes (  
  CountryID int NOT NULL AUTO_INCREMENT,  
  Televotes DECIMAL (3, 0) NOT NULL,  
  PRIMARY KEY(CountryID)  
);
```

The three tables can be seen in the database resultsEurovision in Image 9.



```
mysql> CREATE DATABASE resultsEurovision;  
Query OK, 1 row affected (0.02 sec)  
  
mysql> USE resultsEurovision;  
Database changed
```

Image 5

```

mysql> CREATE TABLE finalists (
->   CountryID int NOT NULL AUTO_INCREMENT,
->   Finalists VARCHAR(255) NOT NULL,
->   PRIMARY KEY(CountryID)
-> );
Query OK, 0 rows affected (0.14 sec)

mysql> INSERT INTO finalists (Finalists) values
-> ('Albania'),
-> ('Azerbaijan'),
-> ('Belgium'),
-> ('Bulgaria'),
-> ('Cyprus'),
-> ('Finland'),
-> ('France'),
-> ('Germany'),
-> ('Greece'),
-> ('Iceland'),
-> ('Israel'),
-> ('Italy'),
-> ('Lithuania'),
-> ('Malta'),
-> ('Moldova'),
-> ('Netherlands'),
-> ('Norway'),
-> ('Portugal'),
-> ('Russia'),
-> ('San Marino'),
-> ('Serbia'),
-> ('Spain'),
-> ('Sweden'),
-> ('Switzerland'),
-> ('Ukraine'),
-> ('United Kingdom');
Query OK, 26 rows affected (0.03 sec)
Records: 26 Duplicates: 0 Warnings: 0

```

Image 6

```

mysql> CREATE TABLE juryvotes (
->   CountryID int NOT NULL AUTO_INCREMENT,
->   Juryvotes DECIMAL(3, 0) NOT NULL,
->   PRIMARY KEY(CountryID)
-> );
Query OK, 0 rows affected (0.14 sec)

mysql> INSERT INTO juryvotes (Juryvotes) values
-> (22),
-> (32),
-> (71),
-> (140),
-> (50),
-> (83),
-> (248),
-> (3),
-> (91),
-> (198),
-> (73),
-> (206),
-> (55),
-> (208),
-> (53),
-> (11),
-> (15),
-> (126),
-> (104),
-> (37),
-> (20),
-> (6),
-> (46),
-> (267),
-> (97),
-> (0);
Query OK, 26 rows affected (0.02 sec)
Records: 26 Duplicates: 0 Warnings: 0

```

Image 7


```
mysql> CREATE TABLE televotes (
->   CountryID int NOT NULL AUTO_INCREMENT,
->   Televotes DECIMAL (3, 0) NOT NULL,
->   PRIMARY KEY(CountryID)
-> );
Query OK, 0 rows affected (0.14 sec)

mysql> INSERT INTO televotes (Televotes) values
-> (35),
-> (33),
-> (3),
-> (30),
-> (44),
-> (218),
-> (251),
-> (0),
-> (79),
-> (180),
-> (20),
-> (318),
-> (165),
-> (47),
-> (62),
-> (0),
-> (60),
-> (27),
-> (100),
-> (13),
-> (82),
-> (0),
-> (63),
-> (165),
-> (267),
-> (0);
Query OK, 26 rows affected (0.01 sec)
Records: 26  Duplicates: 0  Warnings: 0
```

Image 8

```
mysql> USE resultsEurovision;
Database changed
mysql> SHOW TABLES;
+-----+
| Tables_in_resultsEurovision |
+-----+
| finalists                    |
| juryvotes                   |
| televotes                    |
+-----+
3 rows in set (0.00 sec)
```

Image 9

3.2 Entry of the instance data

The dataset which was cleaned is present in the database: the finalist countries are present, the sum of the televotes and the juryvotes are present as well. Compared to the original dataset from the two CSV files, there is obviously a significant drop in the number of columns present, because the level of detail of the original datasets was too complex for the scope of the current project.

The data was inserted using INSERT INTO commands for each of the tables, as follows:

- For the finalists table, as seen in Image 6, the command was:

```
INSERT INTO finalists (Finalists) values
('Albania'),
('Azerbaijan'),
```

('Belgium'),
('Bulgaria'),
('Cyprus'),
('Finland'),
('France'),
('Germany'),
('Greece'),
('Iceland'),
('Israel'),
('Italy'),
('Lithuania'),
('Malta'),
('Moldova'),
('Netherlands'),
('Norway'),
('Portugal'),
('Russia'),
('San Marino'),
('Serbia'),
('Spain'),
('Sweden'),
('Switzerland'),
('Ukraine'),
('United Kingdom');

- For the juryvotes table, as seen in Image 7, the command was:

INSERT INTO juryvotes (Juryvotes) values

(22),
(32),
(71),
(140),
(50),
(83),
(248),
(3),
(91),
(198),
(73),
(206),
(55),
(208),
(53),
(11),

(15),
(126),
(104),
(37),
(20),
(6),
(46),
(267),
(97),
(0);

- For the televotes table, as seen in Image 8, the command was:

INSERT INTO televotes (Televotes) values

(35),
(33),
(3),
(30),
(44),
(218),
(251),
(0),
(79),
(180),
(20),
(318),
(165),
(47),
(62),
(0),
(60),
(27),
(100),
(13),
(82),
(0),
(63),
(165),
(267),
(0);

In conclusion, a significant part of the data was inserted therefore manually, needed for the INSERT INTO command's structure.

3.3. Reflection on how well the database reflects the data

The database is a fair reflection of the dataset used, in its cleaned version. The three tables combine the Finalists, Juryvotes and Televotes in a comprehensive way and it makes use of the sum of the Juryvotes and the Televotes. As mentioned prior, this is not going to further details as in the original datasets coming from the two CSV files, which contained more details in terms of knowing how each country voted. This would have been interesting for a more complex analysis that was determining how countries voted in specific and potentially analyzing their motivations. With the current database, we cannot make a detailed comparison on the differences between how the countries voted in the televotes and the juryvotes. However, a surface of that can be seen in summing the results: we can notice that they don't coincide in establishing the final results of the competition and that the rankings were different in the perspective of the professional jury and the public. This surface can constitute a further motivation for a more detailed analysis on this topic, using the original datasets from the two CSV files.

3.4 SQL commands that answer questions from section 1.3

The finalists of the Eurovision 2021 competition are established using the commands that CREATE the table finalists and the INSERT INTO. They were mentioned previously in the 3.1 and 3.2 sections of the present report and can be seen in Image 6.

In terms of showing the juryvotes in descending order, this is made possible by the following command, which selects the Finalists column from finalists table and the Juryvotes column from the juryvotes table and it joins them together, in descending order:

```
SELECT f.Finalists, j.Juryvotes FROM finalists AS f
JOIN juryvotes AS j ON j.CountryID = f.CountryID
ORDER BY Juryvotes DESC;
```

In terms of showing the televotes in a descending order, this is made possible by the following command, which selects the Finalists column from finalists table and the Televotes column from the televotes table and it joins them together, in descending order:

```
SELECT f.Finalists, j.Televotes FROM finalists AS f
JOIN televotes AS j ON j.CountryID = f.CountryID
ORDER BY Televotes DESC;
```

In terms of showing the overall results of the competition, with both the juryvotes and the televotes added in order to determine the sum of them and therefore the final ranking, in a descending order, there is the following query, which selects the Finalists column from finalists table, the Juryvotes from the juryvotes table and the Televotes column from the televotes table, it makes a sum out of Juryvotes and Televotes called Total, and it joins them together, in descending order:

```
SELECT f.Finalists, g.Juryvotes, j.Televotes, (j.televotes + g.juryvotes) AS Total
```

```
FROM finalists AS f
JOIN juryvotes AS g on g.CountryID = f.CountryID
JOIN televotes AS j ON j.CountryID = f.CountryID
ORDER BY Total DESC;
```

Seeing the data displayed in this manner can show that the rankings have different outcomes: the juryvote established a different ranking for the competition compared to the one which was determined by the public votes. Their sum, however, determined a final ranking which differs as well from both the juryvotes and the televotes, to a certain proportion. These aspects would be interesting to be investigated in a further analysis.

4. Simple web application

4.1 Description of the application

The application has a Navigation bar containing links to:

- The Finalists page
- The Juryvotes page
- The Televotes page
- The Total results page

The application was built both on the local server and in the Coursera lab. On the local server, styling elements were kept, but on the Coursera lab only the html page is displayed, probably given the restrictions of the environment and for example, installing Bootstrap using an external link. I have added screenshots from both environments, for reference, in the 4.2 section.

Regarding the Finalists page, we can notice in Image 10 (from the local server) and Image 11 (from the Coursera lab) that it displays the Finalists column from the finalists table, in alphabetical order. This is to show which are the countries that qualified in the final part of the competition.

In the Juryvotes page, as seen in Image 12 (from the local server) and Image 13 (from the Coursera lab), we can notice the column from the finalists table with the countries that participated in the competition, ordered in descending order based on the juryvotes they've obtained. The Juryvotes column comes from the juryvotes table.

In the Televotes page, as seen in Image 14 (from the local server) and Image 15 (from the Coursera lab), we can notice the column from the finalists table with the countries that participated in the competition, ordered in descending order based on the televotes they've obtained. The Televotes column comes from the televotes table.

Regarding the Total votes page, as seen in Image 16 (from the local server) and Image 17 (from the Coursera lab), we can notice the column from the finalists table with the countries that participated in the competition, ordered in descending order based on the sum of the televotes and juryvotes they've obtained. The Televotes column comes from the televotes table and the Juryvotes column comes from the juryvotes table.

We can notice that these pages answer the questions mentioned initially in section 1.3.

4.2 Screenshots of main screens from the application

Image 10

Finalists	Juryvotes	Televotes	Total Votes
<h2>Eurovision Finalists</h2> <p>Here is a complete list of the Eurovision Finalists.</p>			
Finalists			
Albania			
Azerbaijan			
Belgium			
Bulgaria			
Cyprus			
Finland			
France			
Germany			
Greece			
Iceland			
Israel			
Italy			
Lithuania			
Malta			
Moldova			
Netherlands			
Norway			
Portugal			
Russia			
San Marino			
Serbia			
Spain			
Sweden			
Switzerland			
Ukraine			
United Kingdom			

Image 11

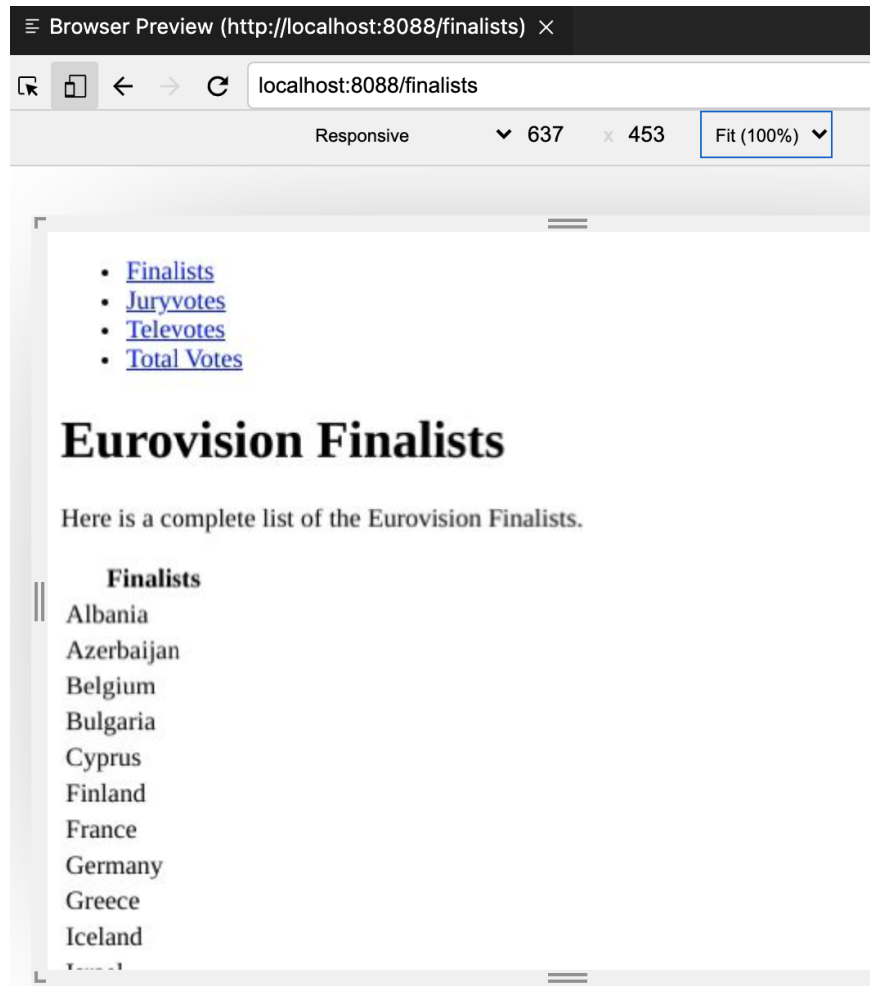


Image 12

Finalists Juryvotes Televotes Total Votes			
Eurovision Juryvotes			
Here is a complete list of the Eurovision Juryvotes.			
Country	Juryvotes		
Switzerland	267		
France	248		
Malta	208		
Italy	206		
Iceland	198		
Bulgaria	140		
Portugal	126		
Russia	104		
Ukraine	97		
Greece	91		
Finland	83		
Israel	73		
Belgium	71		
Lithuania	55		
Moldova	53		
Cyprus	50		
Sweden	46		
San Marino	37		
Azerbaijan	32		
Albania	22		
Serbia	20		
Norway	15		
Netherlands	11		
Spain	6		
Germany	3		
United Kingdom	0		

Image 13

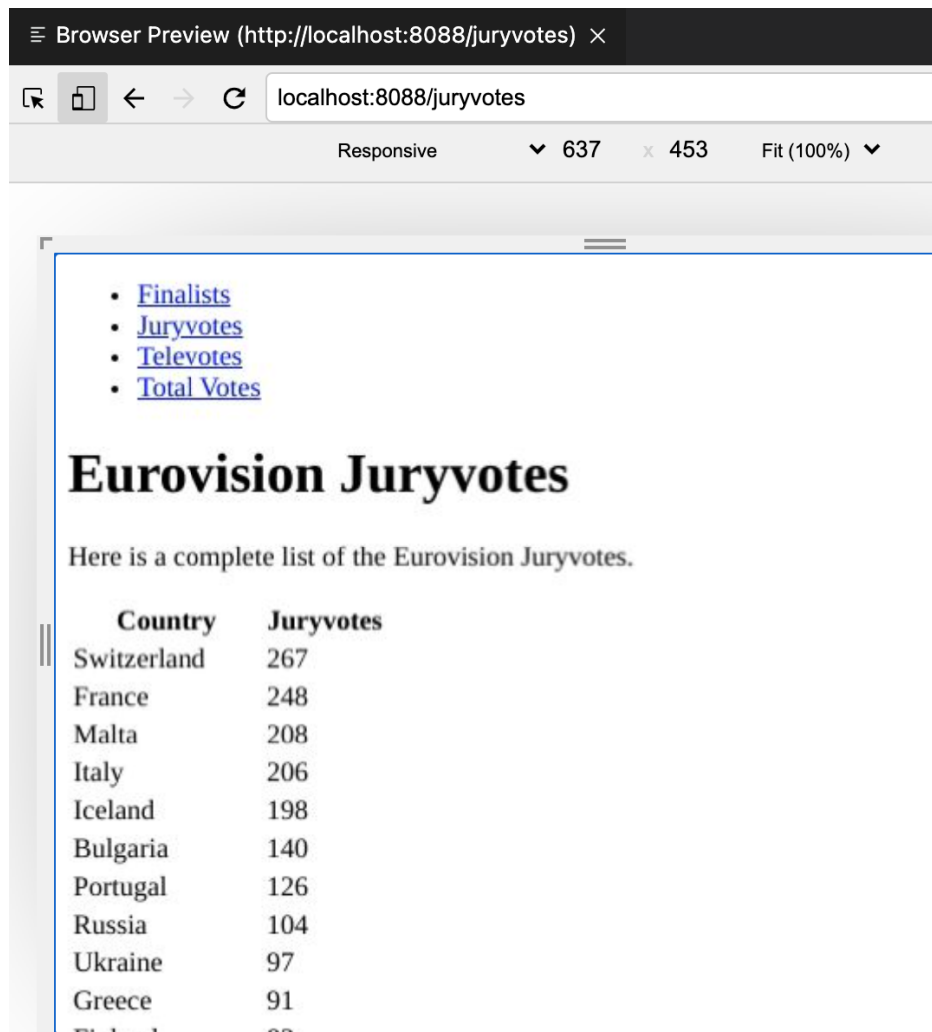


Image 14

Finalists Juryvotes Televotes Total Votes			
Eurovision Televotes			
Here is a complete list of the Eurovision Televotes.			
Country	Televotes		
Italy	318		
Ukraine	267		
France	251		
Finland	218		
Iceland	180		
Lithuania	165		
Switzerland	165		
Russia	100		
Serbia	82		
Greece	79		
Sweden	63		
Moldova	62		
Norway	60		
Malta	47		
Cyprus	44		
Albania	35		
Azerbaijan	33		
Bulgaria	30		
Portugal	27		
Israel	20		
San Marino	13		
Belgium	3		
Germany	0		
Netherlands	0		
Spain	0		
United Kingdom	0		

Image 15

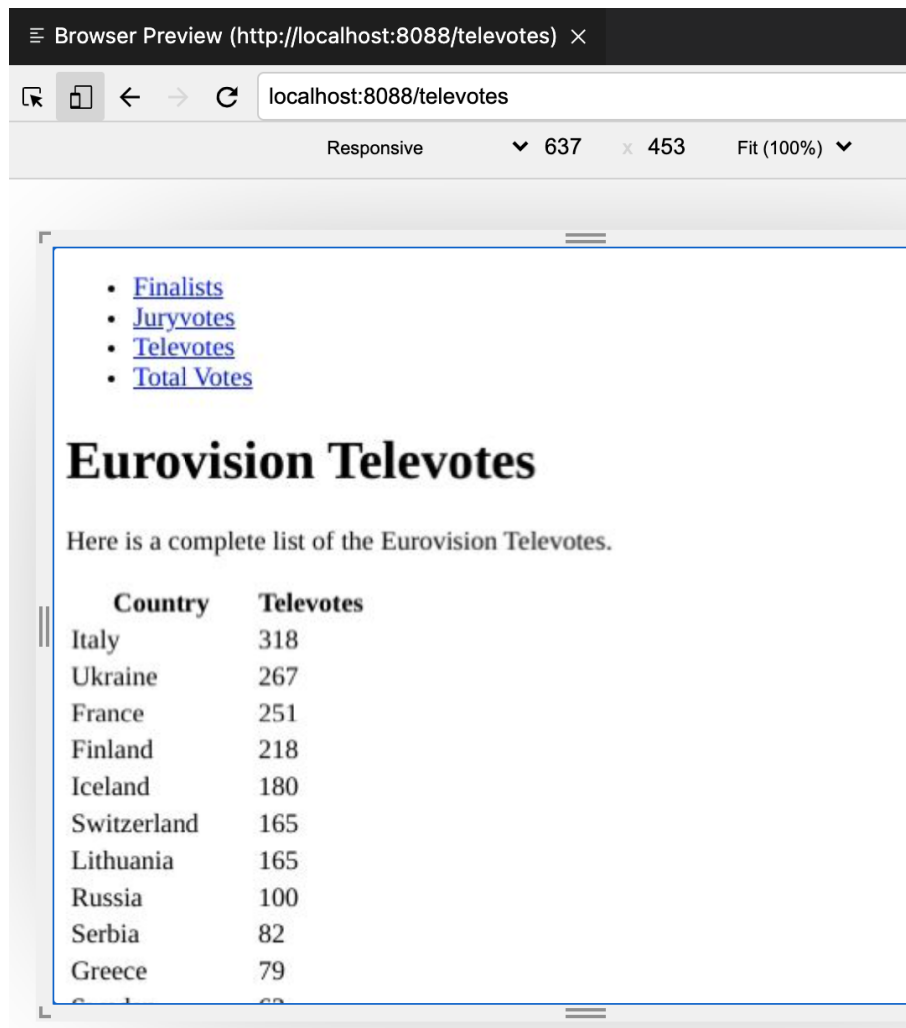


Image 16

Finalists Juryvotes Televotes Total Votes			
Eurovision Total votes			
Here is a complete list of the Eurovision Total Votes.			
Country	Juryvotes	Televotes	Total
Italy	206	318	524
France	248	251	499
Switzerland	267	165	432
Iceland	198	180	378
Ukraine	97	267	364
Finland	83	218	301
Malta	208	47	255
Lithuania	55	165	220
Russia	104	100	204
Bulgaria	140	30	170
Greece	91	79	170
Portugal	126	27	153
Moldova	53	62	115
Sweden	46	63	109
Serbia	20	82	102
Cyprus	50	44	94
Israel	73	20	93
Norway	15	60	75
Belgium	71	3	74
Azerbaijan	32	33	65
Albania	22	35	57
San Marino	37	13	50
Netherlands	11	0	11
Spain	6	0	6
Germany	3	0	3
United Kingdom	0	0	0

Image 17



• [Finalists](#)

• [Juryvotes](#)

• [Televotes](#)

• [Total Votes](#)

Eurovision Total Votes

Here is a complete list of the Eurovision Total Votes.

Country	Juryvotes	Televotes	Total
Italy	206	318	524
France	248	251	499
Switzerland	267	165	432
Iceland	198	180	378
Ukraine	97	267	364
Finland	83	218	301
Malta	208	47	255
Lithuania	55	165	220
Russia	104	100	204
Bulgaria	140	30	170
Greece	91	79	170
Portugal	126	27	153

4.3 Instructions on how to access the Lab

Sharable link:

<https://hub.labs.coursera.org:443/connect/sharednhvgmagn?forceRefresh=false&path=%2F%3Ffolder%3D%2Fhome%2Fcoder%2Fproject>

1. Run node index.js command
2. Write in the browser either one of the followings:
 - a. <localhost:8088/finalists>
 - b. <localhost:8088/juryvotes>
 - c. <localhost:8088/televotes>
 - d. <localhost:8088/totalvotes>

