UNIVERSITÀ DI PISA

LDS Project – Part 3

Group 8:

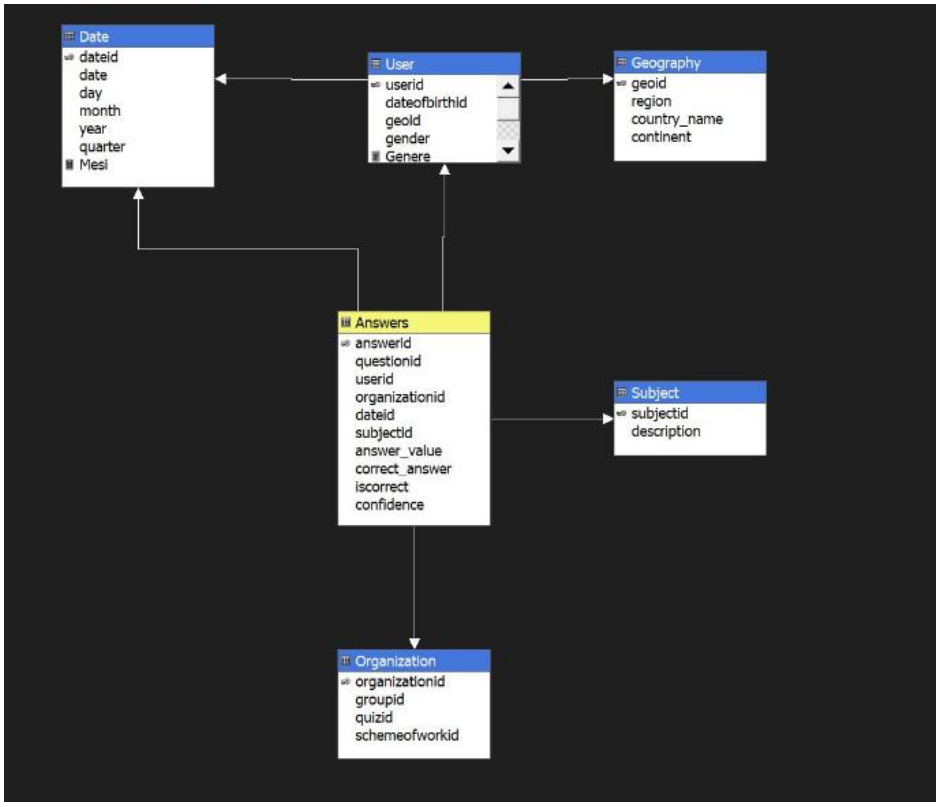
Sabatini Emanuele 637756

Volpi Federico 648069

Assignment 0

> Build a datacube from the data of the tables in your database, defining the appropriate hierarchies for time and geography. Create the needed measures based on the queries you need to answer.

After connecting Visual Studio with the SSAS module to the Data Source of our DB, we have created Data Source views and the proper dimensions with their attributes. The hierarchies for Date and Geography have been set and a cube was created. The Geography hierarchy was created inside of the User dimension. As measures the sum of confidence, the counting of rows and the counting of correct answers were created to answer properly to the next assignments. It is worth mentioning the creation of a new calculated measure in the fact table, as the difference between the counting of Answers and of just the correct ones, that will be useful for the PowerBI tasks. Also, a new attribute has been created in the Date dimension, useful to have the name of months instead of their respective number. It was substituted to the "Month" attribute in the time hierarchy. We have also created a “Genere” attribute in the User dimension as the conversion of the numbers corresponding to each gender as words, to help with PowerBI visualizations.



Assignment 1

Show the total correct answers for each country and the grand total with respect to the continent.

This assignment was solved with the following MDX query.

```
WITH MEMBER Continent_Grand_Total AS
([User].[Geo].currentmember.parent, [Measures].[Correct])

SELECT {[Measures].[Correct], Continent_Grand_Total}  on columns,
nonempty(([User].[Continent].[Continent], [User].[Geo].[Country Name])) on rows
FROM [Group 8 DB]
```

On columns, the total number of correct answers was shown near a grand total referring to the continent, while on rows each country was present. Thus, the result was the total of correct answers for each country near the grand total of the continent in which that country was located.

		Correct	Continent_Grand_Total
Europe	Belgium	12104	275588
Europe	France	15576	275588
Europe	Germany	98976	275588
Europe	Ireland	18608	275588
Europe	Italy	15742	275588
Europe	Spain	51464	275588
Europe	United Kingdom	63118	275588
North America	Canada	19254	39087
North America	United States	19833	39087
Oceania	Australia	14344	27562
Oceania	New Zealand	13218	27562

Assignment 2

Show the total confidence for each year and the running yearly for European students.

```
WITH MEMBER Running_Confidence AS
sum(periodstodate([Date].[Time].[Year], [Date].[Time].currentmember), [Measures].[Total Confidence])

MEMBER Annual_Confidence AS
([Date].[Time].parent.parent, [Measures].[Total Confidence])

SELECT {Running_Confidence, Annual_Confidence} on columns,
nonempty(([Date].[Year].[Year], [Date].[Time].[Mesi])) on rows
FROM [Group 8 DB]
WHERE [User].[Geo].[Continent].&[Europe]
```

We have considered on rows the various months and as columns first of all Running_Confidence, that is the monthly cumulative confidence, that restarts every year. Then, in another column named Annual_Confidence, the total confidence for each year.

		Running_Confidence	Annual_Confidence
2019	January	544225	23586100
2019	February	4043125	23586100
2019	March	7458925	23586100
2019	April	8977775	23586100
2019	May	10328800	23586100
2019	June	11402600	23586100
2019	July	11832900	23586100
2019	August	11912275	23586100
2019	September	14224225	23586100
2019	October	18234150	23586100
2019	November	21754225	23586100
2019	December	23586100	23586100
2020	January	2445550	7651000
2020	February	4483050	7651000
2020	March	6716875	7651000
2020	April	7651000	7651000

Assignment 3

Show the ratio between the total correct answers of each year w.r.t the previous year.

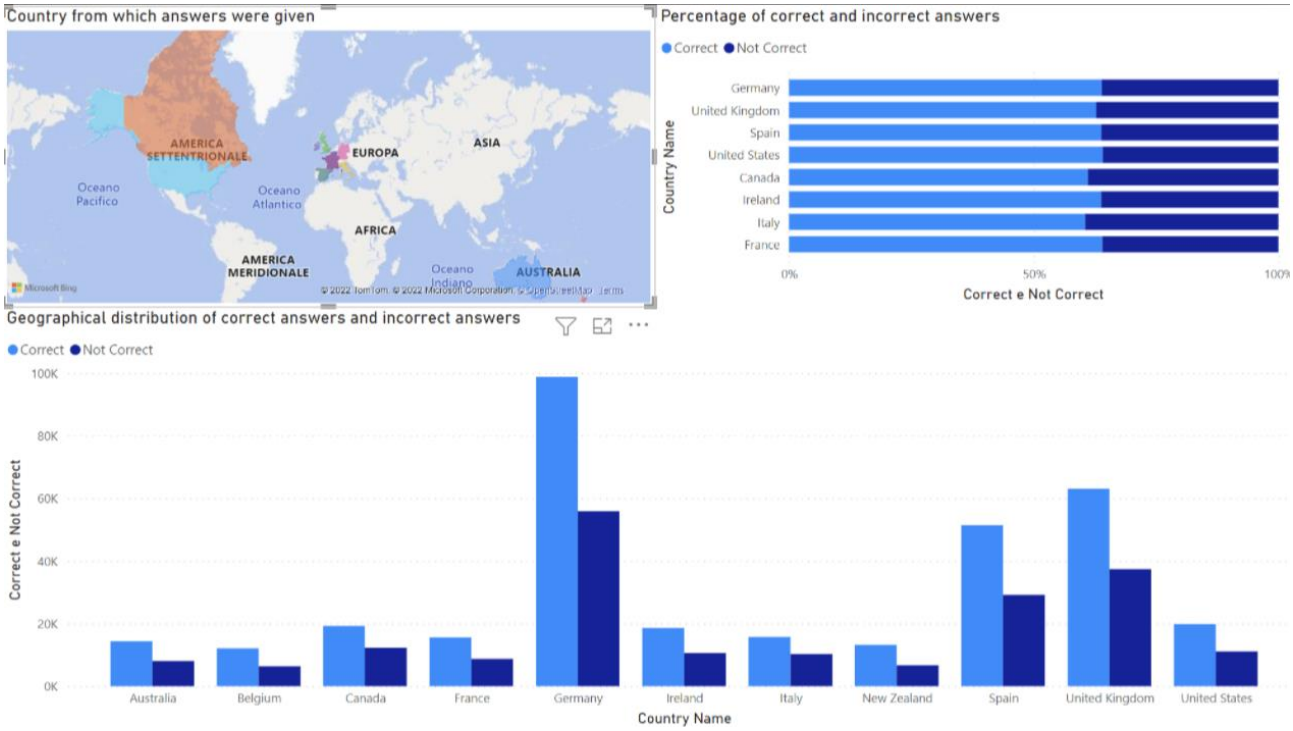
```
WITH MEMBER correct_ratio AS
iif(([Date].[Time].currentmember.lag(1), [Measures].[Correct]) = 0, 1,
    (([Date].[Time].currentmember, [Measures].[Correct]) /
        ([Date].[Time].currentmember.lag(1), [Measures].[Correct]))),
FORMAT_STRING = "Percent"

SELECT correct_ratio on columns,
nonempty([Date].[Time].[Year]) on rows
FROM [Group 8 DB]
```

To solve the assignment, a column correct_ratio was created and using .lag() it is possible to compute the ratio between the total correct answers for the current year, that is the current member of the hierarchy, over the previous year. The “iif” is used in order to manage division by zero. With FORMAT_STRING we set the percentage as format.

	correct_ratio
2019	100.00%
2020	32.91%

Assignment 4

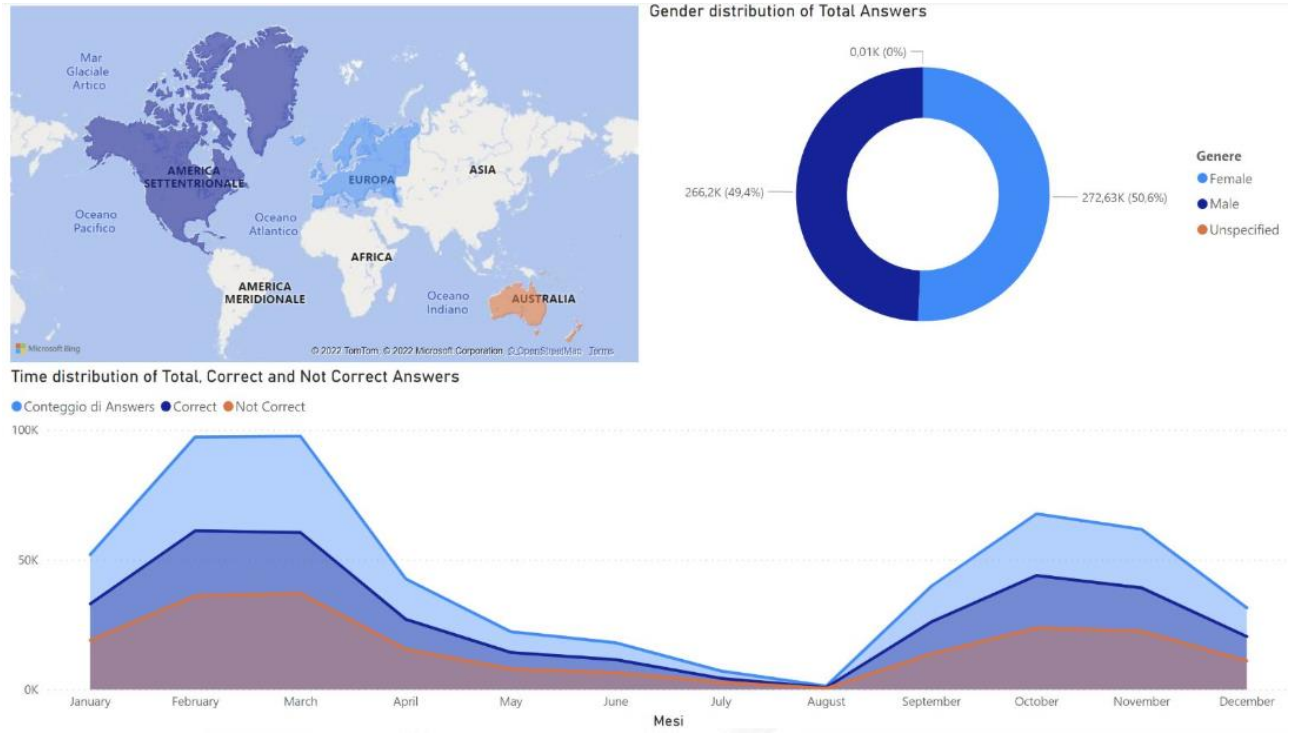
Create a dashboard that shows the geographical distribution of *correct answers* and *incorrect answers*



This assignment was solved using PowerBI. Our Dashboard contained an interactive map with the countries in which the answers were given, an horizontal stacked bar chart to show the percentage of correct answers in the various countries and a bar chart that considers the absolute number of correct and not correct answers.

Assignment 5

Create a plot/dashboard of your choosing, that you deem interesting w.r.t. the data available in your cube



In PowerBI, we have decided to create an interactive dashboard that shows the time distribution of total, correct and not correct answers in their absolute number. It is possible to show the distribution by gender or by continent, in the proper interactive map.