

**ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS**

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
DEPARTMENT OF MANAGEMENT, SCIENCE & TECHNOLOGY
MSc BUSINESS ANALYTICS

“Data Warehousing & Visualization”

Full Name: ATHANASIOS ALEXANDRIS
Register Number: p2822202
&
Full Name: KRISTJANA KOLAJ
Register Number: p2822213

ATHENS, 2022

Table of Contents

1. Business Case.....	page 3
2. Report's Targets.....	page 3
3. Data Cleaning.....	page 4
4. Data Import.....	page 9
5. Data Cube Deployment.....	page 15
6. Business Analysis.....	page 18

Business Case

Glassdoor produces reports based on the data collected from its users, on topics including work-life balance, CEO pay-ratios, lists of the best office places and cultures, and the accuracy of corporate job searching maxims. Data from Glassdoor has also been used by outside sources to produce estimates on the effects of salary trends and changes on corporate revenues. Glassdoor also puts the conclusions of its research of other companies towards its own company policies.

Our dataset represents reviews on UK Firms by employees. We use data as for example the date of the review, the job name, the job location, the status of the reviewers, and the reviews. Reviews are divided in s sub-categories Career Opportunities, Comp & Benefits, Culture & Values, Senior Management, and Work/Life Balance. In addition, employees added recommendations on the firm, the CEO, and the outlook.

The Great Resignation, employees quitting, job seekers demanding better benefits and work-life balance... the past year was a game-changer as we continued to recover from the effects of the pandemic. But that didn't stop today's biggest company giants from supporting their employees and rising up as some of the best companies to work for. See the Top 10 Companies to work for.

Report's Targets

- 1) Top 10 Firms for an employee to work for based on average rating
- 2) Top 10 Firms for an employee to work form depending on which aspect wants to focus (1. Work Life Balance, 2. Culture and Values, 3. Composition & Benefits, 4. Career Opportunities, 5. Diversity Inclusion, 6. Senior Management and 7. Overall Rating.
- 3) Top 10 Firms sorted by recommendations on the firm, the CEO, and the outlook

1.Dataset

The dataset we chose to work on, is called 'Glassdoor Job Reviews' and is provided in the following URL: <https://www.kaggle.com/datasets/davidgauthier/glassdoor-job-reviews>.

It is large dataset of Glassdoor reviews platform, that contains job descriptions and rankings among various criteria such as work-life balance, income, culture, etc. for various cities globally, for the period 31/1/2008-8/6/2021.

Glassdoor is one of the fastest growing jobs and recruiting sites. Glassdoor produces reports based upon the data collected from its users, on topics including work–life balance, CEO pay-ratios, lists of the best office places and cultures, and the accuracy of corporate job searching maxims. Data from Glassdoor has also been used by outside sources to produce estimates on the effects of salary trends and changes on corporate revenues.

The dataset contains 838566 rows and 18 columns. The columns correspond to the date of the review, the job name, the job location, the status of the reviewers, and the reviews. Reviews are divided in sub-categories Career Opportunities, Comp & Benefits, Culture &

Values, Senior Management, and Work/Life Balance. In addition, employees can add recommendations on the firm, the CEO, and the outlook.

Ranking for the recommendation of the firm, CEO approval, and outlook are allocated categories v, r, x, and o, with the following meanings:

v - Positive, r - Mild, x - Negative, o - No opinion

Also there are columns with comments for the advantages and the disadvantages

Below are listed the columns of the dataset:

1. firm
2. date_review
3. job_title
4. current
5. location
6. overall_rating
7. work_life_balance
8. culture_values
9. diversity_inclusion
10. career_opp
11. comp_benefits
12. senior_mgmt
13. recommend
14. ceo_approv
15. outlook
16. headline
17. pros
18. cons

Cleaning

We used R-studio to have a quick insight of our data set. Using the summary function, we analyzed each column to find the variable type, minimum/maximum and mean values.

```
> summary(dff8[,2:21])
  firm      date_review      job_title      current      tenure      city
Length:838566 Length:838566 Length:838566 Length:838566 Length:838566 Length:838566
Class :character Class :character Class :character Class :character Class :character Class :character
Mode  :character Mode  :character Mode  :character Mode  :character Mode  :character Mode  :character

  location      overall_rating      work_life_balance      culture_values      diversity_inclusion      career_opp      comp_benefits
Length:838566 Min. :1.000 Min. :1.00 Min. :1.00 Min. :1 Min. :1.00 Min. :1.0
Class :character 1st Qu.:3.000 1st Qu.:2.00 1st Qu.:3.00 1st Qu.:3 1st Qu.:3.00 1st Qu.:3.0
Mode :character Median :4.000 Median :4.00 Median :4.00 Median :4 Median :4.00 Median :3.0
Mean :3.656 Mean :3.38 Mean :3.59 Mean :4 Mean :3.46 Mean :3.4
3rd Qu.:5.000 3rd Qu.:4.00 3rd Qu.:5.00 3rd Qu.:5 3rd Qu.:5.00 3rd Qu.:4.0
Max. :5.000 Max. :5.00 Max. :5.00 Max. :5 Max. :5.00 Max. :5.0
NA's :149894 NA's :191373 NA's :702500 NA's :147501 NA's :150082

  senior_mgmt      recommend      ceo_approv      outlook      headline      pros
Min. :1.00 Length:838566 Length:838566 Length:838566 Length:838566 Length:838566
1st Qu.:2.00 Class :character Class :character Class :character Class :character Class :character
Median :3.00 Mode :character Mode :character Mode :character Mode :character Mode :character
Mean :3.18
3rd Qu.:4.00
Max. :5.00
NA's :155876

  cons
Length:838566
Class :character
Mode :character
```

First we observed that column 'current' contains information for the employee, regarding the type of working relationship (employee, contractor, intern, temporary employee), current or former status, and the tenure of the employee in the company (more than 1 year, less than 1 year etc.)

So we decided, that for analytics purposes it might be more convenient to split this column to three columns for each one category described above.

The separation was made with the 'separate' function of 'tidyr' package.

R function:

```
dff2<-separate(dff1,col=current,sep="," , into= c("current","tenure"))
```

```
dff9<-separate(dff8,col=current,sep=" " , into= c("current_former","classification"))
```

Below are the screenshots of before and after the transformation.

Before

After

current
Current Employee
Current Employee, more than 1 year
Current Employee, less than 1 year
Current Employee
Current Employee, more than 1 year
Current Employee, less than 1 year
Former Employee
Current Employee, more than 5 years
Former Employee, more than 1 year
Former Employee, more than 3 years
Current Employee, less than 1 year

current_former	classification	tenure
Current	Employee	NA
Current	Employee	more than 1 year
Current	Employee	less than 1 year
Current	Employee	NA
Current	Employee	more than 1 year
Current	Employee	less than 1 year
Former	Employee	NA
Current	Employee	more than 5 years
Former	Employee	more than 1 year
Former	Employee	more than 3 years
Current	Employee	less than 1 year

Next, we observed that in column location, there were values with double words. So we identified the columns that this situation was observed and we fixed this.

Below are the screenshots of before and after the transformation.

Before

location
Bromsgrove, England, England
Bromsgrove, England, England
Bromsgrove, England, England
Bromsgrove, England, England

location
Bromsgrove, England
Bromsgrove, England
Bromsgrove, England
Bromsgrove, England
Century City, CA
Santa Ana, CA
San Juan Capistrano, CA

Another difficulty that we faced, regarding the column 'location', is that the values of the columns have not equal number of words, so it was not easy to separate this column to other columns based on their hierarchy.

location
Wisbech, England
Blyth, North East England, England
Manchester, England
Woking, England
Crawley, West Sussex, South East England, England

However, we observed that always the first word refers to the city. So we decided to add another column 'city' that refers to the first word of the column 'location'.

city	location
Bromsgrove	Bromsgrove, England
Bromsgrove	Bromsgrove, England
Bromsgrove	Bromsgrove, England
Bromsgrove	Bromsgrove, England
Century City	Century City, CA
Santa Ana	Santa Ana, CA
San Juan Capistrano	San Juan Capistrano, CA

Finally, we changed blanks with N/A values for the variables we are planning to use as Dimensions on our Fact Table later.

```
dff8[df8==""]<-NA
dff8[df8==" "]<-NA
```

job_title	current	tenure	city
NA	Current Employee	NA	NA
Office Administrator	Current Employee	more than 1 year	Bromsgrove
Office Administrator	Current Employee	less than 1 year	Bromsgrove
NA	Current Employee	NA	NA
Office Administrator	Current Employee	more than 1 year	Bromsgrove
Office Administrator	Current Employee	less than 1 year	NA
IFA	Former Employee	NA	Bromsgrove
Anonymous Employee	Current Employee	more than 5 years	Century City
Anonymous Employee	Former Employee	more than 1 year	NA
Technician	Former Employee	more than 3 years	Santa Ana

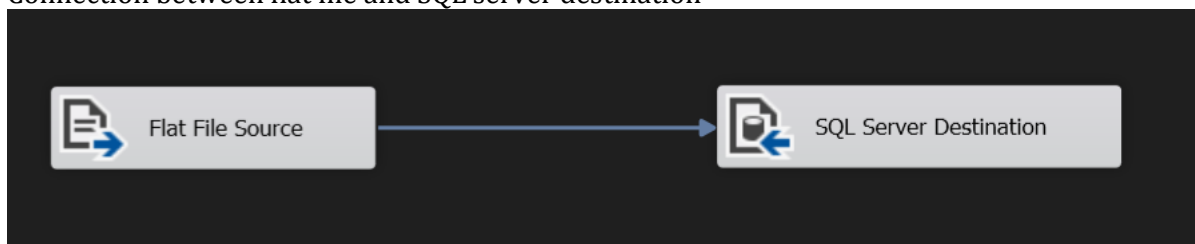
Our numerical measure columns, will contain blanks instead of NAs, in order to be set as null values in the database.

```
dff8[,9:15][is.na(dff8[,9:15])]<- ""
```

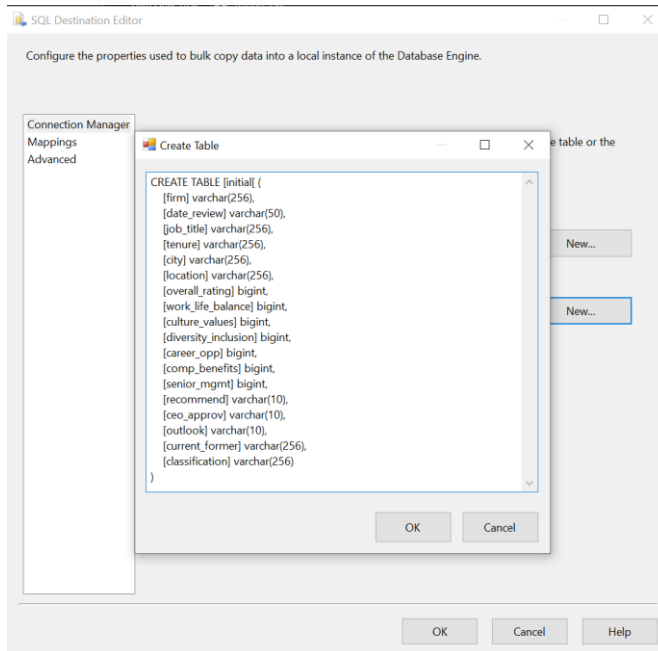
Data Import

First, we created a relational database in SQL Server called “job_reviews”. Then, in Visual Studio, we created a Data Flow Task called “Import CSV” which includes a Flat-file Source and an SQL Server Destination. The Flat-file Source was imported with the cleaned CSV file of our dataset. We connected SQL Server Destination to the Flat-file Source in order to connect our CSV directly to the database “job_reviews” in SQL Server and used it to create a new table called “initial” in our database. By creating table “staging”, we declared the data types of our columns.

Connection between flat file and SQL server destination



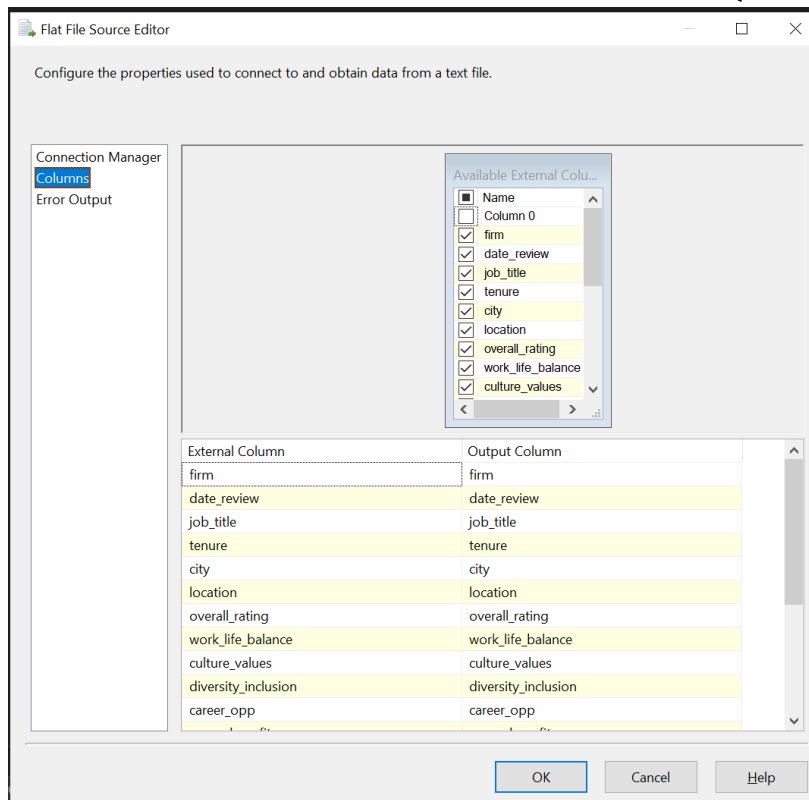
Creation of Table “initial”



Column Name	Data Type	Allow Nulls
firm	varchar(256)	<input checked="" type="checkbox"/>
date_review	varchar(50)	<input checked="" type="checkbox"/>
job_title	varchar(256)	<input checked="" type="checkbox"/>
current_former	varchar(256)	<input checked="" type="checkbox"/>
classification	varchar(256)	<input checked="" type="checkbox"/>
tenure	varchar(256)	<input checked="" type="checkbox"/>
city	varchar(256)	<input checked="" type="checkbox"/>
location	varchar(256)	<input checked="" type="checkbox"/>
overall_rating	bigint	<input checked="" type="checkbox"/>
work_life_balance	bigint	<input checked="" type="checkbox"/>
culture_values	bigint	<input checked="" type="checkbox"/>
diversity_inclusion	bigint	<input checked="" type="checkbox"/>
career_opp	bigint	<input checked="" type="checkbox"/>
comp_benefits	bigint	<input checked="" type="checkbox"/>
senior_mgmt	bigint	<input checked="" type="checkbox"/>
recommend	varchar(10)	<input checked="" type="checkbox"/>
ceo_approv	varchar(10)	<input checked="" type="checkbox"/>
outlook	varchar(10)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Finally, we chose the columns we will use in our report, connected the Flat-file Source to the SQL Server Destination and imported the CSV file to the “initial” table. We decided not to choose the first column ‘Column0’ which involves indexes of rows and it was created automatically in the export of the csv file from R, and the three last columns that are related to comments, as we did not plan to do analysis of comments.

Selection of Columns and Connection of Flat-file Source to SQL Server Destination

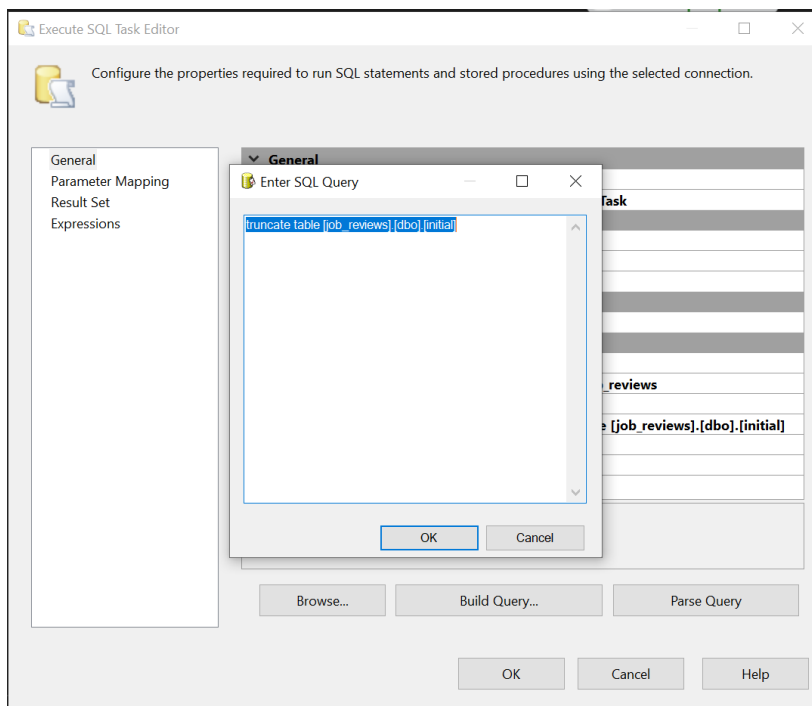


Furthermore, we connected the “Import CSV” process, with an SQL Execute Task, that would truncate our “staging” table each time we execute the data flow.

Connection of “Truncate Staging” to “Import CSV”



Truncate Query Execution



Afterwards, we defined the dimensions and measures for the fact table. After consideration, we decided that the following columns will be used as dimensions:

- firm
- date_review
- job_title
- current_former
- classification
- tenure
- city
- location
- recommend
- ceo_approv
- outlook

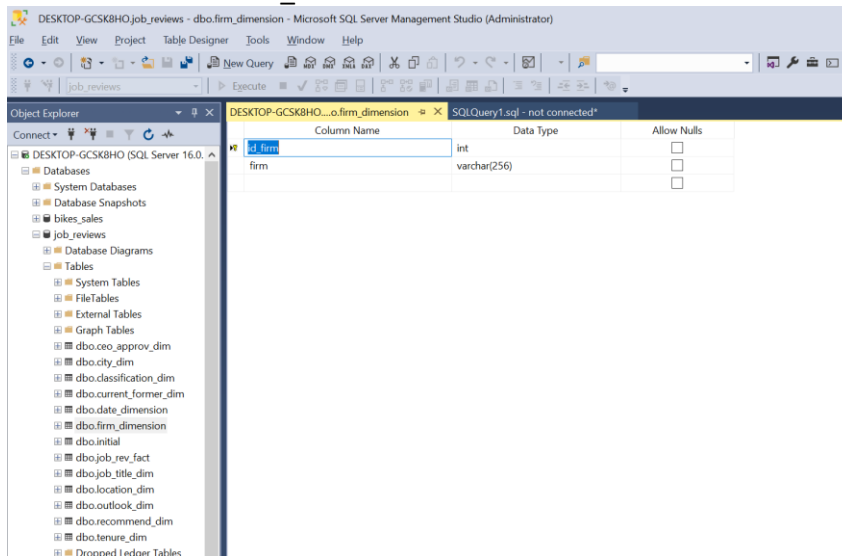
and the following will be used as measures:

- overall_rating
- work_life_balance

- culture_values
- diversity_inclusion
- career_opp
- comp_benefits
- senior_mgmt

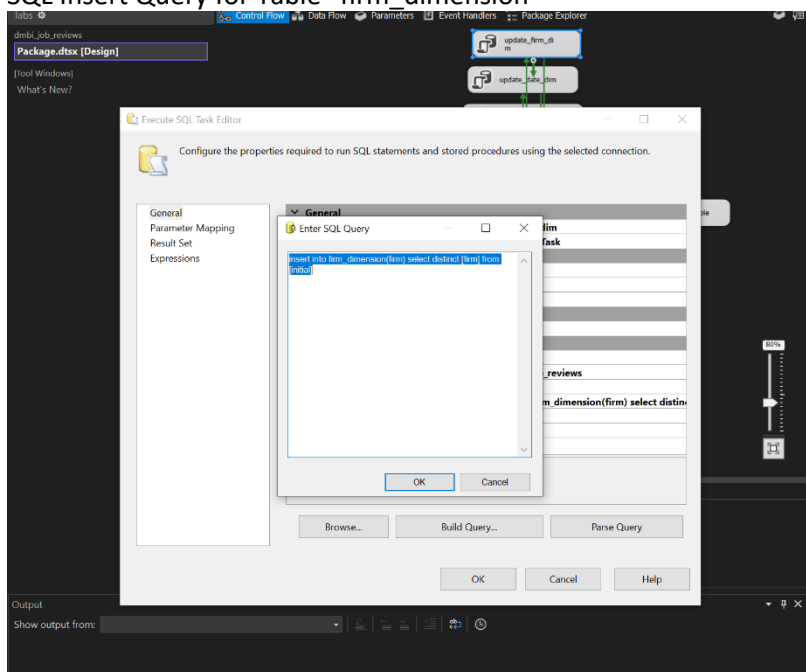
To implement that, we created a new table for each dimension. In the following image we demonstrate the construction of a dimension table with a primary key and a column to insert the values. The same procedure was followed for all the other Dimensions.

Creation of Table “firm_dimension”



Afterwards, in Visual Studio, we created an Execute SQL Task for each dimension which contains the SQL Query that inserts values into our dimension tables and connected each one of them with the “Import CSV”.

SQL Insert Query for Table “firm_dimension”



Finally, we created the fact table called “job_rev_fact” which contains as foreign keys the primary keys of each dimension and the 7 measures we described before.

Fact Table Design and SQL Insert Query

The top screenshot shows the design of the `job_rev_fact` table in Microsoft SQL Server Enterprise Manager. The table has the following columns and data types:

Column Name	Data Type	Allow Nulls
firm	int	<input type="checkbox"/>
date_review	int	<input type="checkbox"/>
job_title	int	<input type="checkbox"/>
current_former	int	<input type="checkbox"/>
classification	int	<input type="checkbox"/>
tenure	int	<input type="checkbox"/>
city	int	<input type="checkbox"/>
location	int	<input type="checkbox"/>
recommend	int	<input type="checkbox"/>
ceo_approv	int	<input type="checkbox"/>
outlook	int	<input type="checkbox"/>
overall_rating	int	<input checked="" type="checkbox"/>
work_life_balance	int	<input checked="" type="checkbox"/>
culture_values	int	<input checked="" type="checkbox"/>
diversity_inclusion	int	<input checked="" type="checkbox"/>
career_opp	int	<input checked="" type="checkbox"/>
comp_benefits	int	<input checked="" type="checkbox"/>
senior_mgmt	int	<input checked="" type="checkbox"/>

The bottom screenshot shows the `Execute SQL Task Editor` in Visual Studio, displaying the following SQL query:

```
INSERT INTO job_rev_fact
SELECT [firm_dimension].[id_firm] AS firm, [date_dimension].[id_date] AS [date_review], [job_title_dim].[id_job_title] AS [job_title],
[current_former_dim].[id_current_former] AS [current_former], [classification_dim].[id_classification] AS [classification],
[tenure_dim].[id_tenure] AS [tenure], [city_dim].[id_city] AS [city], [location_dim].[id_location] AS [location],
[recommend_dim].[id_recommend] AS [recommend], [ceo_approv_dim].[id_ceo_approv] AS [ceo_approv],
[outlook_dim].[id_outlook] AS [outlook],
[initial].[overall_rating], [initial].[work_life_balance], [initial].[culture_values], [initial].[diversity_inclusion], [initial].[career_opp],
[initial].[comp_benefits], [initial].[senior_mgmt]
FROM [job_reviews].[dbo].[initial]
INNER JOIN [firm_dimension] ON [initial].[firm] = [firm_dimension].[firm]
INNER JOIN [date_dimension] ON [initial].[date_review] = [date_dimension].[date]
INNER JOIN [job_title_dim] ON [initial].[job_title] = [job_title_dim].[job_title]
INNER JOIN [current_former_dim] ON [initial].[current_former] = [current_former_dim].[current_former]
INNER JOIN [classification_dim] ON [initial].[classification] = [classification_dim].[classification]
INNER JOIN [tenure_dim] ON [initial].[tenure] = [tenure_dim].[tenure]
INNER JOIN [city_dim] ON [initial].[city] = [city_dim].[city]
INNER JOIN [location_dim] ON [initial].[location] = [location_dim].[location]
INNER JOIN [recommend_dim] ON [initial].[recommend] = [recommend_dim].[recommend]
INNER JOIN [ceo_approv_dim] ON [initial].[ceo_approv] = [ceo_approv_dim].[ceo_approv]
INNER JOIN [outlook_dim] ON [initial].[outlook] = [outlook_dim].[outlook]
```

Afterwards, in Visual Studio, we created an Execute SQL Task for fact table which contains the SQL Query that inserts values into our fact table and connects it with each one dimension’s SQL Execute Task. The final running structure of our control flow in Visual Studio is the following:

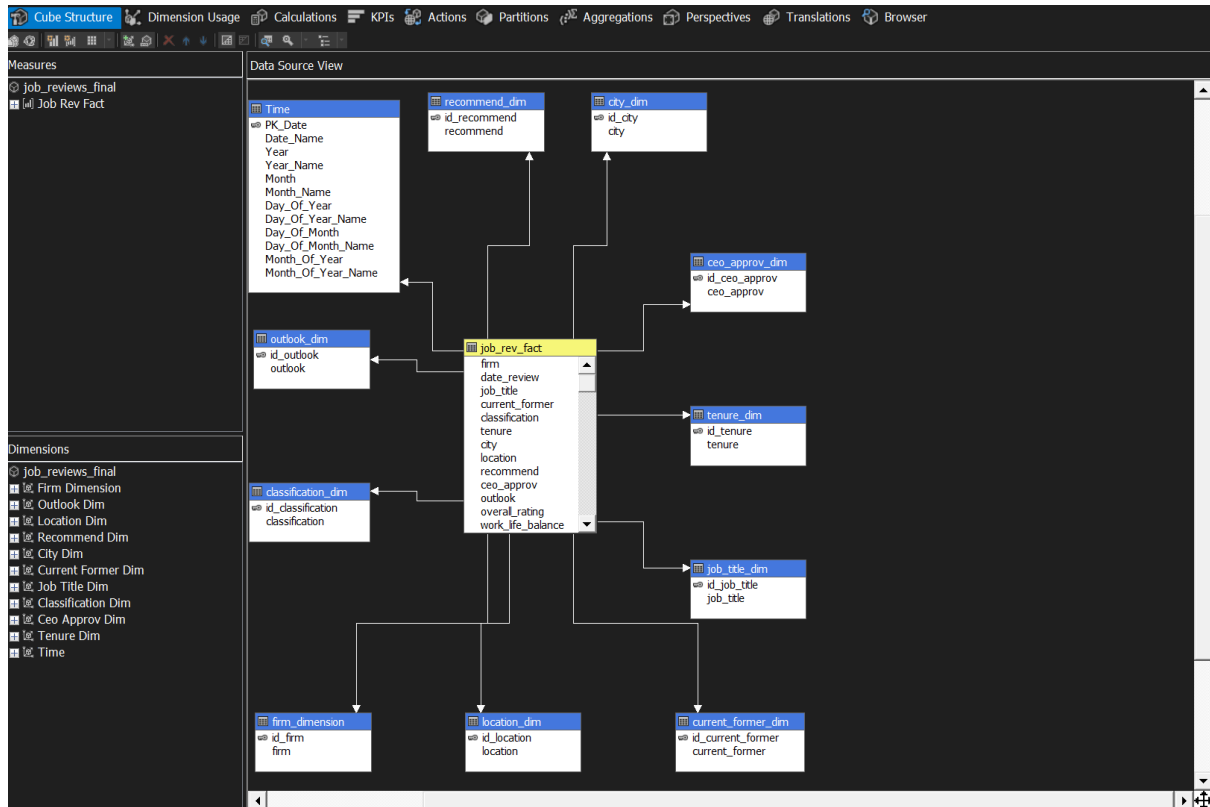
Connection of “job_rev_fact” to the other Execute SQL Tasks:

SQLQuery4.sql - D:\SK8HO\thanos (67) - * × SQLQuery3.sql - D:\SK8HO\thanos (66) - * × SQLQuery2.sql - not connected * SQLQuery1.sql - not connected																																																																																																																																																																																																																																																																																																																																																																																																					
/***** Script for SelectTopRows command from SSMS *****/																																																																																																																																																																																																																																																																																																																																																																																																					
<pre> SELECT FROM [job_reviews].[dbo].[job_rev_fact] </pre>																																																																																																																																																																																																																																																																																																																																																																																																					
100 %																																																																																																																																																																																																																																																																																																																																																																																																					
<table border="1"> <thead> <tr> <th>firm</th><th>job_review</th><th>job_title</th><th>current_former</th><th>classification</th><th>tenure</th><th>city</th><th>location</th><th>recommend</th><th>ceo_approval</th><th>outlook</th><th>overall_rating</th><th>work_life_balance</th><th>culture_values</th><th>diversity_inclusion</th><th>career_opportunities</th><th>comp_benefits</th><th>senior_mgmt</th></tr> </thead> <tbody> <tr><td>10270</td><td>2021-05-14 00:00:00.000</td><td>1727548</td><td>57</td><td>170</td><td>203</td><td>350315</td><td>412279</td><td>83</td><td>111</td><td>108</td><td>4</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1727548</td><td>57</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>83</td><td>111</td><td>108</td><td>5</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1735274</td><td>57</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>83</td><td>111</td><td>108</td><td>5</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1735274</td><td>57</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>82</td><td>111</td><td>105</td><td>3</td><td>3</td><td>3</td><td>3</td><td>5</td><td>3</td><td>4</td></tr> <tr><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1735274</td><td>57</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>83</td><td>111</td><td>108</td><td>5</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1735274</td><td>57</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>82</td><td>111</td><td>105</td><td>3</td><td>3</td><td>3</td><td>3</td><td>5</td><td>3</td><td>4</td></tr> <tr><td>10480</td><td>2019-10-17 00:00:00.000</td><td>1734299</td><td>57</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>82</td><td>111</td><td>105</td><td>3</td><td>4</td><td>4</td><td>NULL</td><td>3</td><td>3</td><td>3</td></tr> <tr><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1747688</td><td>58</td><td>170</td><td>203</td><td>350315</td><td>412279</td><td>83</td><td>111</td><td>108</td><td>4</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1750031</td><td>57</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>83</td><td>111</td><td>108</td><td>4</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1750678</td><td>57</td><td>170</td><td>198</td><td>350879</td><td>412974</td><td>83</td><td>111</td><td>108</td><td>4</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>11</td><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1770143</td><td>58</td><td>170</td><td>203</td><td>363889</td><td>417858</td><td>83</td><td>111</td><td>108</td><td>3</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>12</td><td>10480</td><td>2021-05-14 00:00:00.000</td><td>1770274</td><td>57</td><td>170</td><td>198</td><td>360369</td><td>413606</td><td>84</td><td>111</td><td>107</td><td>1</td><td>1</td><td>1</td><td>5</td><td>1</td><td>5</td></tr> <tr><td>13</td><td>10450</td><td>2019-10-16 00:00:00.000</td><td>1744534</td><td>57</td><td>170</td><td>200</td><td>355937</td><td>408111</td><td>82</td><td>111</td><td>108</td><td>4</td><td>4</td><td>5</td><td>NULL</td><td>4</td><td>2</td></tr> <tr><td>14</td><td>10480</td><td>2021-05-15 00:00:00.000</td><td>1724047</td><td>58</td><td>170</td><td>203</td><td>358007</td><td>410658</td><td>83</td><td>111</td><td>108</td><td>3</td><td>1</td><td>3</td><td>3</td><td>3</td><td>3</td></tr> <tr><td>15</td><td>10480</td><td>2021-05-15 00:00:00.000</td><td>1725767</td><td>58</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>83</td><td>111</td><td>108</td><td>4</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>16</td><td>10450</td><td>2019-10-16 00:00:00.000</td><td>1725767</td><td>57</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>83</td><td>111</td><td>108</td><td>4</td><td>5</td><td>4</td><td>NULL</td><td>3</td><td>3</td></tr> <tr><td>17</td><td>10480</td><td>2021-05-15 00:00:00.000</td><td>1728652</td><td>58</td><td>170</td><td>197</td><td>360369</td><td>413606</td><td>84</td><td>111</td><td>105</td><td>3</td><td>1</td><td>3</td><td>5</td><td>4</td><td>3</td></tr> <tr><td>18</td><td>10480</td><td>2021-05-15 00:00:00.000</td><td>1735274</td><td>58</td><td>170</td><td>200</td><td>361841</td><td>415375</td><td>83</td><td>111</td><td>108</td><td>4</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>19</td><td>10450</td><td>2019-10-15 00:00:00.000</td><td>1725767</td><td>58</td><td>170</td><td>203</td><td>360369</td><td>413606</td><td>83</td><td>111</td><td>108</td><td>3</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>20</td><td>10450</td><td>2019-10-15 00:00:00.000</td><td>1725670</td><td>57</td><td>170</td><td>198</td><td>360369</td><td>413606</td><td>83</td><td>111</td><td>108</td><td>2</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td><td>NULL</td></tr> <tr><td>21</td><td>10480</td><td>2021-05-15 00:00:00.000</td><td>1744302</td><td>57</td><td>170</td><td>200</td><td>361138</td><td>414541</td><td>83</td><td>111</td></tr></tbody></table>	firm	job_review	job_title	current_former	classification	tenure	city	location	recommend	ceo_approval	outlook	overall_rating	work_life_balance	culture_values	diversity_inclusion	career_opportunities	comp_benefits	senior_mgmt	10270	2021-05-14 00:00:00.000	1727548	57	170	203	350315	412279	83	111	108	4	NULL	NULL	NULL	NULL	NULL	NULL	10480	2021-05-14 00:00:00.000	1727548	57	170	203	360369	413606	83	111	108	5	NULL	NULL	NULL	NULL	NULL	NULL	10480	2021-05-14 00:00:00.000	1735274	57	170	203	360369	413606	83	111	108	5	NULL	NULL	NULL	NULL	NULL	NULL	10480	2021-05-14 00:00:00.000	1735274	57	170	203	360369	413606	82	111	105	3	3	3	3	5	3	4	10480	2021-05-14 00:00:00.000	1735274	57	170	203	360369	413606	83	111	108	5	NULL	NULL	NULL	NULL	NULL	NULL	10480	2021-05-14 00:00:00.000	1735274	57	170	203	360369	413606	82	111	105	3	3	3	3	5	3	4	10480	2019-10-17 00:00:00.000	1734299	57	170	203	360369	413606	82	111	105	3	4	4	NULL	3	3	3	10480	2021-05-14 00:00:00.000	1747688	58	170	203	350315	412279	83	111	108	4	NULL	NULL	NULL	NULL	NULL	NULL	10480	2021-05-14 00:00:00.000	1750031	57	170	203	360369	413606	83	111	108	4	NULL	NULL	NULL	NULL	NULL	NULL	10480	2021-05-14 00:00:00.000	1750678	57	170	198	350879	412974	83	111	108	4	NULL	NULL	NULL	NULL	NULL	NULL	11	10480	2021-05-14 00:00:00.000	1770143	58	170	203	363889	417858	83	111	108	3	NULL	NULL	NULL	NULL	NULL	12	10480	2021-05-14 00:00:00.000	1770274	57	170	198	360369	413606	84	111	107	1	1	1	5	1	5	13	10450	2019-10-16 00:00:00.000	1744534	57	170	200	355937	408111	82	111	108	4	4	5	NULL	4	2	14	10480	2021-05-15 00:00:00.000	1724047	58	170	203	358007	410658	83	111	108	3	1	3	3	3	3	15	10480	2021-05-15 00:00:00.000	1725767	58	170	203	360369	413606	83	111	108	4	NULL	NULL	NULL	NULL	NULL	16	10450	2019-10-16 00:00:00.000	1725767	57	170	203	360369	413606	83	111	108	4	5	4	NULL	3	3	17	10480	2021-05-15 00:00:00.000	1728652	58	170	197	360369	413606	84	111	105	3	1	3	5	4	3	18	10480	2021-05-15 00:00:00.000	1735274	58	170	200	361841	415375	83	111	108	4	NULL	NULL	NULL	NULL	NULL	19	10450	2019-10-15 00:00:00.000	1725767	58	170	203	360369	413606	83	111	108	3	NULL	NULL	NULL	NULL	NULL	20	10450	2019-10-15 00:00:00.000	1725670	57	170	198	360369	413606	83	111	108	2	NULL	NULL	NULL	NULL	NULL	21	10480	2021-05-15 00:00:00.000	1744302	57	170	200	361138	414541	83	111
firm	job_review	job_title	current_former	classification	tenure	city	location	recommend	ceo_approval	outlook	overall_rating	work_life_balance	culture_values	diversity_inclusion	career_opportunities	comp_benefits	senior_mgmt																																																																																																																																																																																																																																																																																																																																																																																				
10270	2021-05-14 00:00:00.000	1727548	57	170	203	350315	412279	83	111	108	4	NULL	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
10480	2021-05-14 00:00:00.000	1727548	57	170	203	360369	413606	83	111	108	5	NULL	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
10480	2021-05-14 00:00:00.000	1735274	57	170	203	360369	413606	83	111	108	5	NULL	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
10480	2021-05-14 00:00:00.000	1735274	57	170	203	360369	413606	82	111	105	3	3	3	3	5	3	4																																																																																																																																																																																																																																																																																																																																																																																				
10480	2021-05-14 00:00:00.000	1735274	57	170	203	360369	413606	83	111	108	5	NULL	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
10480	2021-05-14 00:00:00.000	1735274	57	170	203	360369	413606	82	111	105	3	3	3	3	5	3	4																																																																																																																																																																																																																																																																																																																																																																																				
10480	2019-10-17 00:00:00.000	1734299	57	170	203	360369	413606	82	111	105	3	4	4	NULL	3	3	3																																																																																																																																																																																																																																																																																																																																																																																				
10480	2021-05-14 00:00:00.000	1747688	58	170	203	350315	412279	83	111	108	4	NULL	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
10480	2021-05-14 00:00:00.000	1750031	57	170	203	360369	413606	83	111	108	4	NULL	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
10480	2021-05-14 00:00:00.000	1750678	57	170	198	350879	412974	83	111	108	4	NULL	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
11	10480	2021-05-14 00:00:00.000	1770143	58	170	203	363889	417858	83	111	108	3	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
12	10480	2021-05-14 00:00:00.000	1770274	57	170	198	360369	413606	84	111	107	1	1	1	5	1	5																																																																																																																																																																																																																																																																																																																																																																																				
13	10450	2019-10-16 00:00:00.000	1744534	57	170	200	355937	408111	82	111	108	4	4	5	NULL	4	2																																																																																																																																																																																																																																																																																																																																																																																				
14	10480	2021-05-15 00:00:00.000	1724047	58	170	203	358007	410658	83	111	108	3	1	3	3	3	3																																																																																																																																																																																																																																																																																																																																																																																				
15	10480	2021-05-15 00:00:00.000	1725767	58	170	203	360369	413606	83	111	108	4	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
16	10450	2019-10-16 00:00:00.000	1725767	57	170	203	360369	413606	83	111	108	4	5	4	NULL	3	3																																																																																																																																																																																																																																																																																																																																																																																				
17	10480	2021-05-15 00:00:00.000	1728652	58	170	197	360369	413606	84	111	105	3	1	3	5	4	3																																																																																																																																																																																																																																																																																																																																																																																				
18	10480	2021-05-15 00:00:00.000	1735274	58	170	200	361841	415375	83	111	108	4	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
19	10450	2019-10-15 00:00:00.000	1725767	58	170	203	360369	413606	83	111	108	3	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
20	10450	2019-10-15 00:00:00.000	1725670	57	170	198	360369	413606	83	111	108	2	NULL	NULL	NULL	NULL	NULL																																																																																																																																																																																																																																																																																																																																																																																				
21	10480	2021-05-15 00:00:00.000	1744302	57	170	200	361138	414541	83	111																																																																																																																																																																																																																																																																																																																																																																																											

Data Cube Deployment

We created a Multidimensional Analysis Services project in SSAS and connected it to our SQL Server relational database to create a cube with the help of Cube Wizard. Our final cube Schema is the following:

Cube Visualization



We matched dimensions and measures of our database to the Cube and calculated some extra measures. Most of our measures represent averages (the averages of our already existing measures), other measure that were used in the calculation of average measures(total sum of measures, count of non empty values etc.) , and percentages of particular values of our dimensions(the percentage of reviewers that replied positive to recommend the firm. These measures are created in order to use them later in our visualizations.

Below are listed some examples of the measures that we created.

Name:

[TOTAL__SUM]

Parent Properties

Parent hierarchy:

Measures

Parent member:

Change

Expression

[Measures].[Carrer Opp]+[Measures].[Comp Benefits]+[Measures].[Culture Values]+[Measures].[Diversity Inclusion]+[Measures].[Overall Rating]+[Measures].[Senior Mgmt]+[Measures].[Work Life Balance]

No issues found

Ln: 1Ch: 196Col: 95SPCCRLF

Additional Properties

Name:

[TOTAL__AVERAGE]

Parent Properties

Parent hierarchy:

Measures

Parent member:

Change

Expression

[Measures].[TOTAL__SUM]/[Measures].[TOTAL_NON_EMPTY_COUNT]

No issues found

Ln: 1Ch: 59SPCCRLF

Additional Properties

Format string:

Visible:

True

Non-empty behavior:

Name:

[TOTAL_NON_EMPTY_COUNT]

Parent Properties

Parent hierarchy:

Measures

Parent member:

Change

Expression

[Measures].[Carrer Opp Count non nulls]+[Measures].[Comp Benefits Count non nulls]+[Measures].[Culture Values Count non nulls]+[Measures].[Diversity Inclusion Count non nulls]+[Measures].[Overall Rating

No issues found

Ln: 1Ch: 308Col: 7SPCCRLF

Additional Properties

Format string:

Visible:

True

Name:
[AVERAGE_WORK_LIFE_BALANCE]

Parent Properties
Parent hierarchy: Measures
Parent member:
Change

Expression
[Measures].[Work Life Balance]/[Measures].[Work Life Balance Count non nulls]
No issues found
Ln: 1 Ch: 78 SPC CRLF

Additional Properties
Format string:
Visible: True

Name:
[AVERAGE CEO APPROV 'x']

Parent Properties
Parent hierarchy: Measures
Parent member:
Change

Expression
[Measures].[COUNT CEO APPROV 'x']/[Measures].[Job Rev Fact Count]
No issues found
Ln: 1 Ch: 66 SPC CRLF

Additional Properties
Format string: "Percent"

Name:
[COUNT CEO APPROV 'x']

Parent Properties
Parent hierarchy: Measures
Parent member:
Change

Expression
([Measures].[Job Rev Fact Count],[Ceo Approv Dim].[Ceo Approv].[x])
No issues found
Ln: 1 Ch: 68 SPC CRLF

Additional Properties
Format string:

It is worth to refer to the creation of the measure 'weighted average per firm'.

In order we have a ranking between our firms, we could use the total rating of each firm, but this is not that correct by the mathematical scope . For example if a firm has total rating 4.5 and total number of reviews 1.000, it is ranked higher from a firm that has 4.1 and total number of reviews 100.000 . So, in order to have a more accurate view of our data we created a new measure that into account the combination of Total rating for each firm, and also the total count of reviews for each firm.

The formula that we used to calculate this measure is described below:

$$W.Rate = \frac{rating * count + w * r}{count + w}$$

w = average count of reviews per company

r = median of the column

link: <https://stackoverflow.com/a/50476254>

Name:

Parent Properties

Parent hierarchy:

Parent member:

Change

Expression

☒ No issues found

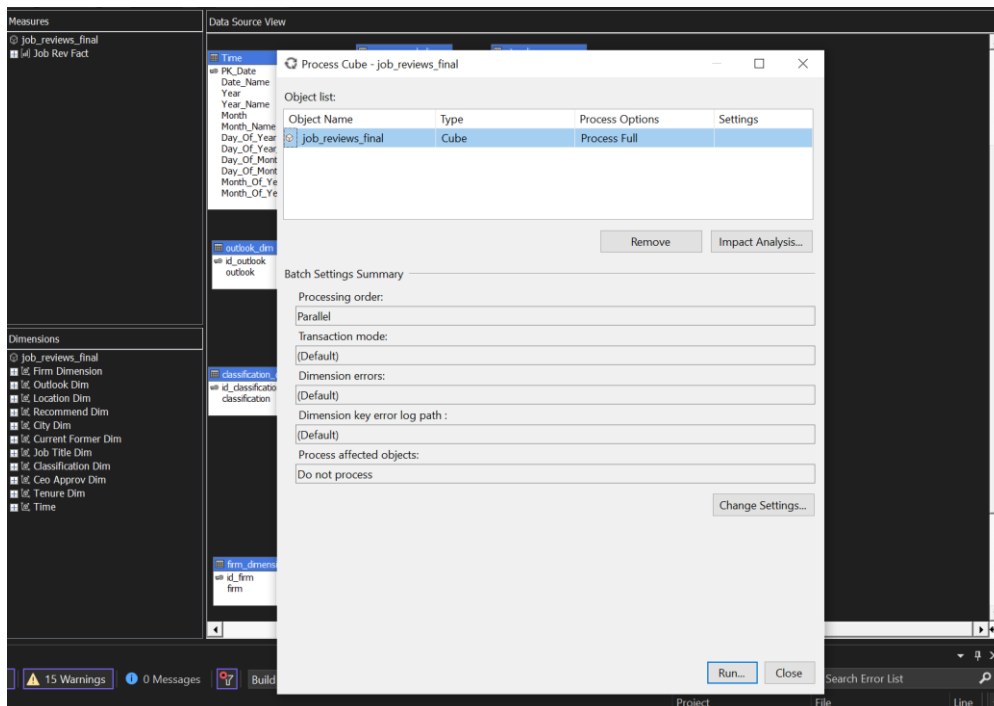
Ln: 1 Ch: 125 Col: 34 SPC CRLF

Additional Properties

Format string:

Visible:

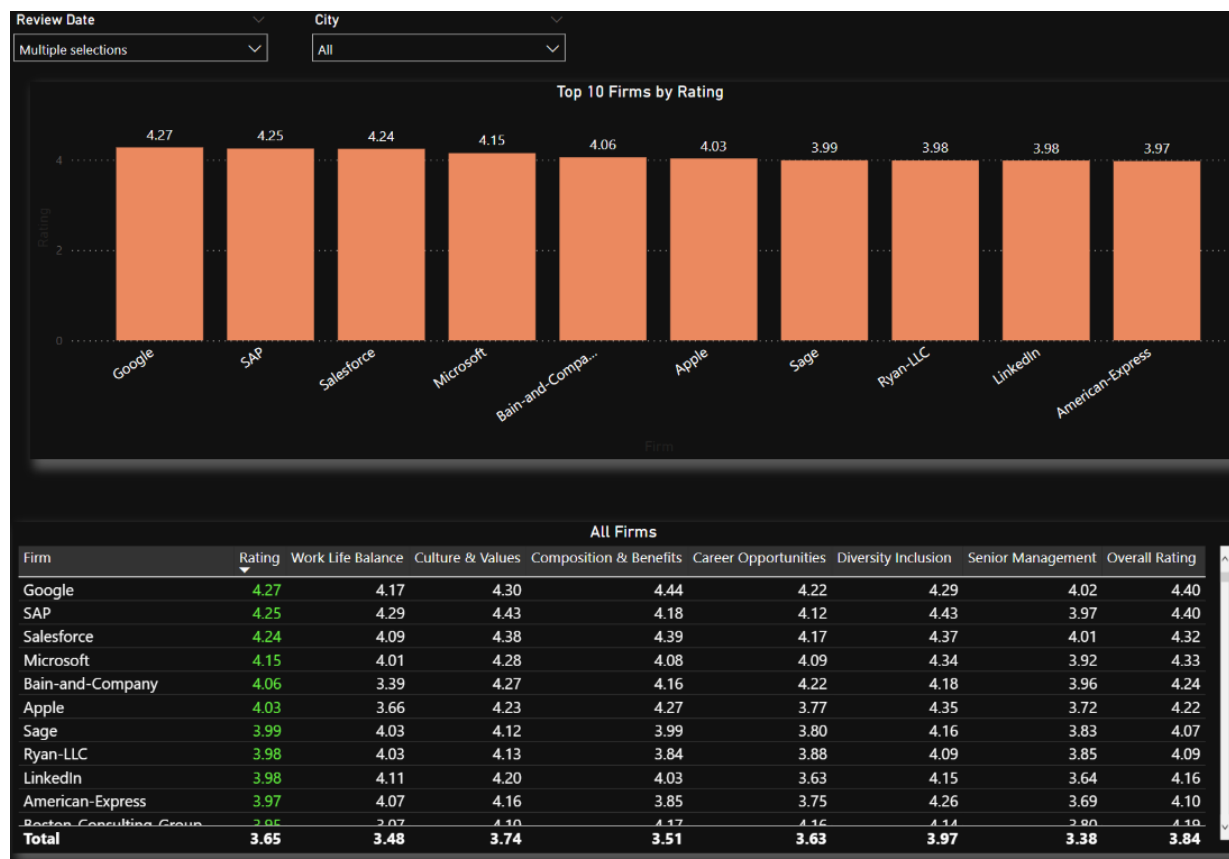
Finally, we process the cube and we connected it to Power BI to visualize our data.



Business Analysis

1. Using Power Bi, we created a main visualization about the Top 10 places for an employee to work, for 2020-2021. On the top, we created a stacked column chart which presents the average rating of Top 10 firms, based on employees' reviews. With a rating scale from 1-5, the Top 10 Firms are Google (4.27), Sap (4.25), Salesforce (4.24), Microsoft (4.15), Bain & Company (4.06), Apple (4.03), Sage (3.99), Ryan-LLC, LinkedIn (3.98) and American-Express (3.97).

While no formula can capture the idiosyncrasies of these companies and the indicative ways they motivate employees, our analysis and derivation of the final ratings was based on 7 parameters of common themes. Specifically, we listed a table showing the Total Ratings, which result from below parameters: 1. Work Life Balance, 2. Culture and Values, 3. Composition & Benefits, 4. Career Opportunities, 5. Diversity Inclusion, 6. Senior Management and 7. Overall Rating.



2. Given the fact that Job seekers may want to focus on a specific aspect of interest, we considered necessary to create stacked column charts presenting the Top 10 Firms, based on each parameter of interest. In the examples below we can notice that examining in terms of 1.Senior Management Google comes first, in terms of 2.Career Opportunities McKinsey and Company comes first, in terms of 3.Competition & Benefits Google comes first, in terms of 4.Culture & Values Sap comes first, in terms of 5.Overall Rating Google comes first, in terms of 6.Work Life Balance Sap comes first and in terms of 7.Diversity Inclusion Sap comes first.

By doing this analysis it was observed that there are companies that stood out when examining by certain aspect of interest but are not in the Top 10 companies based on Total Rating. Such companies are: 1. Sage, 2. McKinsey and Company, 3. Boston Consulting Group, 4. PwC, 5. Deloitte, 6. EY, 7. Bain and Company, 8. University of Michigan, 9. Thomson Reuters, 10. Indeed, 11. Booking-com, 12. Mastercard, 13. Nando-s UK-and-IRE, 14. FARFETCH and 15. Age-UK-The-National-Charity. These companies are colored with yellow.



Review Date
Multiple selections

Senior Management

Diversity Inclusion

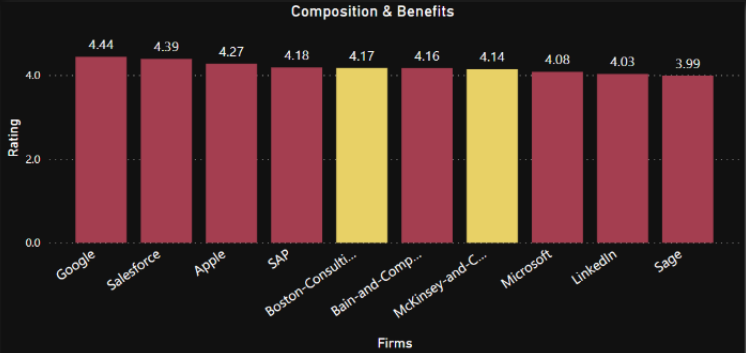
Career Opportunities

Composition &
Benefits

Culture & Values

Overall Rating

Work Life Balance



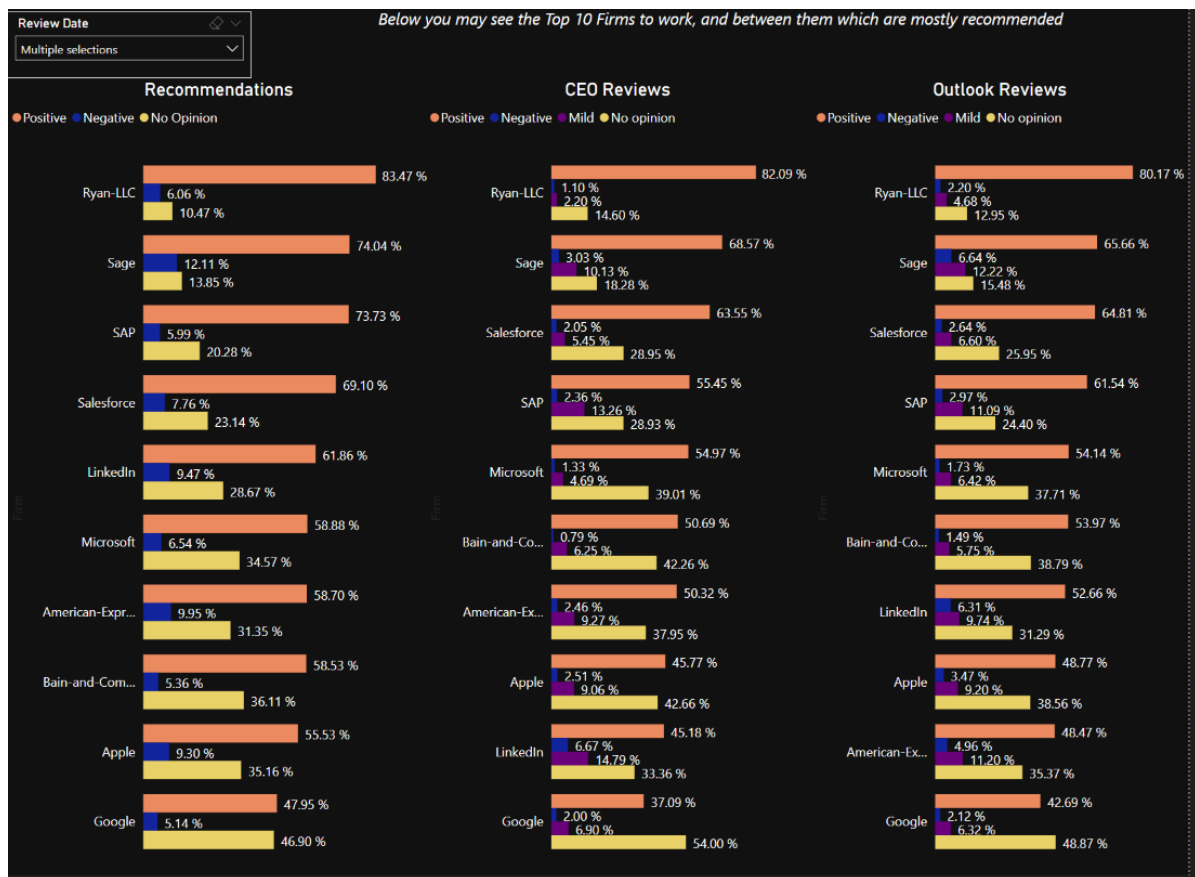
Firm	Rating
Google	4.27
SAP	4.25
Salesforce	4.24
Microsoft	4.15
Bain-and-Company	4.06
Apple	4.03
Sage	3.99
Ryan-LLC	3.98
LinkedIn	3.98
American-Express	3.97



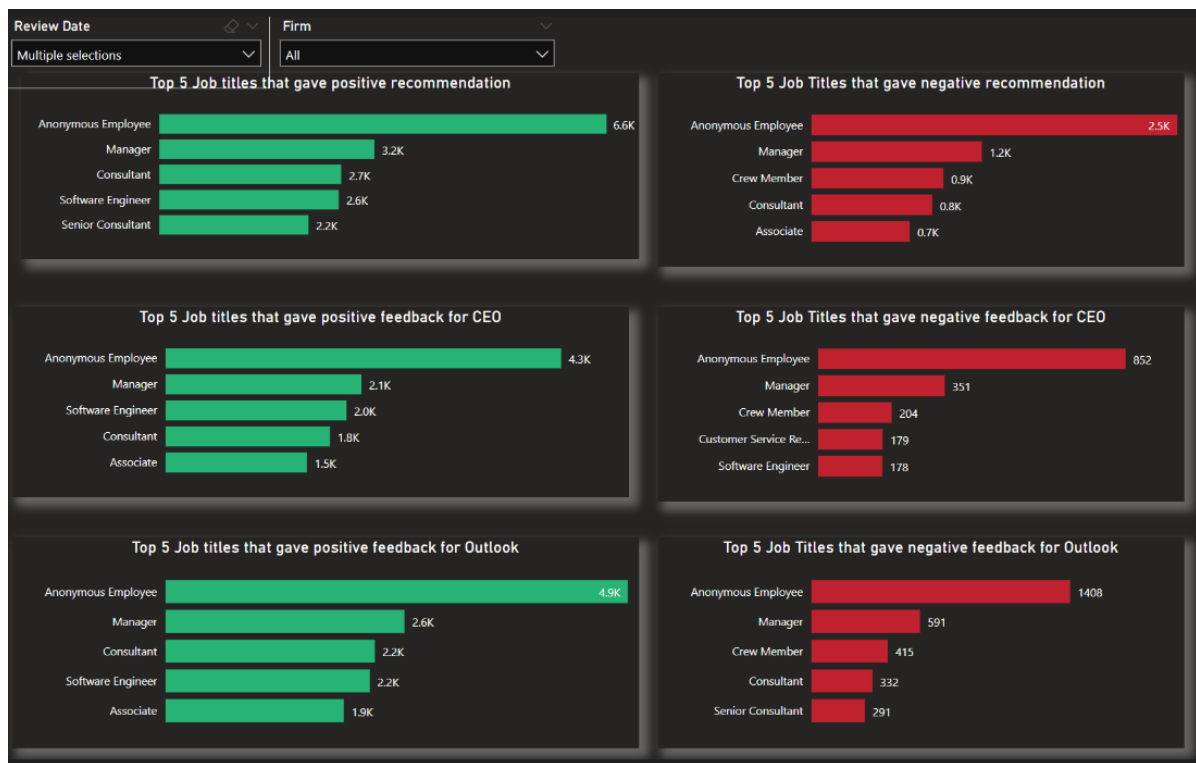


- As it is not enough to analyze some specific aspects to capture the success of a company, we considered it necessary to focus on some other quality data provided by the Reviewers. Employees can add positive, mild, negative and no opinion recommendations on the firm, the CEO, and the Outlook.

We analyzed the Top 10 Firms, based on these three aspects in order to observe the percentage of positive, mild, negative and no opinion reviews. The results are presented below: 1) Ryan-LLC has the most positive reviews in recommendations on the Firm, the CEO and also the Outlook, while Sage has the most negative in recommendations on the Firm and Outlook and finally LinkedIn has the most negative reviews in CEO. One could say that this makes sense and correlates with the Total Rating results, as these companies are in the bottom 4 of the Top 10 Firms.



- We considered it necessary to examine if there is any correlation regarding the Job role of each reviewer and the way they recommend the firm, the CEO and the Outlook. Most negative and positive comments come from Anonymous Employees. Then we observed that most positive reviews come from managers, software engineers, consultants, and associates. While negative come from managers, crew members, consultants, associates, customer service representatives, software engineers and senior consultants.



5. Finally we considered it necessary to do an analysis based on the year in order to see the trend of Total rating for the Top 10 Firms, from 2019-2021. We noticed that the line with the highest rates was in 2020. Therefore we see a small reduction in rates in 2021. Nevertheless, these rates are still high. The past year was a game-changer as we continued to recover from the effects of the pandemic. But that didn't stop today's biggest companies alike from supporting their employees and rising up as some of the best companies to work for.

Year

All

Firm

All

