Start-ups in Europe

Shreeya Namboori
18128947

April 12, 2019

Abstract

Startups play an important part in the economic growth of the country. Although Europe has become a startup hub, there are still many countries in Europe that have lower entrepreneurship rate than others. There is a way to increase the entrepreneurship rate of countries in Europe if we know the factors that are affecting it. The results of one BI query shows that the country having higher corporate tax have higher female unemployment and lower entrepreneurship rate compared to the country with lower corporate tax that has lower female unemployment and higher entrepreneurship rate. No relation could be found out between corporate tax, percentage of startups in growth stage and no. of fastest growing private companies in Europe (INC 5000 Europe).So we cannot say that the countries with lower corporate tax have higher number of fastest growing private companies or higher percentage of startups in growth stage .Another result from the last BI query shows that countries having higher number of fastest growing private companies (INC 5000 Europe) have higher GDP. The aim of this DWBI project is to find factors that affects entrepreneurship rate and see how the entrepreneurship in a country affect its GDP and unemployment.

1 Introduction

This project is about Start-ups in Europe. Start-ups play a vital role in the economic growth of a country. They contribute to the job creation in the market more than an established businesses (Parvinrouh Doddering 2017).Europe is emerging as a start-up hub with many of its start-up earning a global recognition. Europe that was previously lagging behind Silicon Valley and Asia is rapidly coming into its own as a start-up hub (Smale 2018) .Europe is a home to many Unicorn companies (A unicorn is a privately held start-up company (in finance)that is valued at over 1 billion Dollar). It also has an INC 5000 Europe list which is basically the list of approximately 5000 privately held and fastest growing companies all across Europe. The aim of this DWBI project is to find out the why some countries seem to have more entrepreneurship rate than the others and if the entrepreneurship seems to have any impact on various factors like GDP and Unemployment rate of a Country. The three Business Intelligence (BI) Queries for my project are -

1. Does the country with higher corporate tax has lower entrepreneurship rate and higher female unemployment than the country with lower corporate tax?
2. Do the countries with lower corporate tax have more number of fastest growing private companies (INC 5000 Europe list) and more percentage of startups in growth rate?

3. Do the countries that have more number of fastest growing private companies (INC 5000 Europe list) have higher GDP?

2 Data Sources

I have used 6 datasets for this project:

2.1 Source 1: Statista

This dataset gives information about entrepreneurship activity rate of 20 European countries for the year 2018.After cleaning the data we get 20 rows of data(excluding column name) with two columns for Country and Entrepreneurship rate.Link- `https://www.statista.com/statistics/315502/percentage-of-population-involved-in-business-start-ups-in-europe`

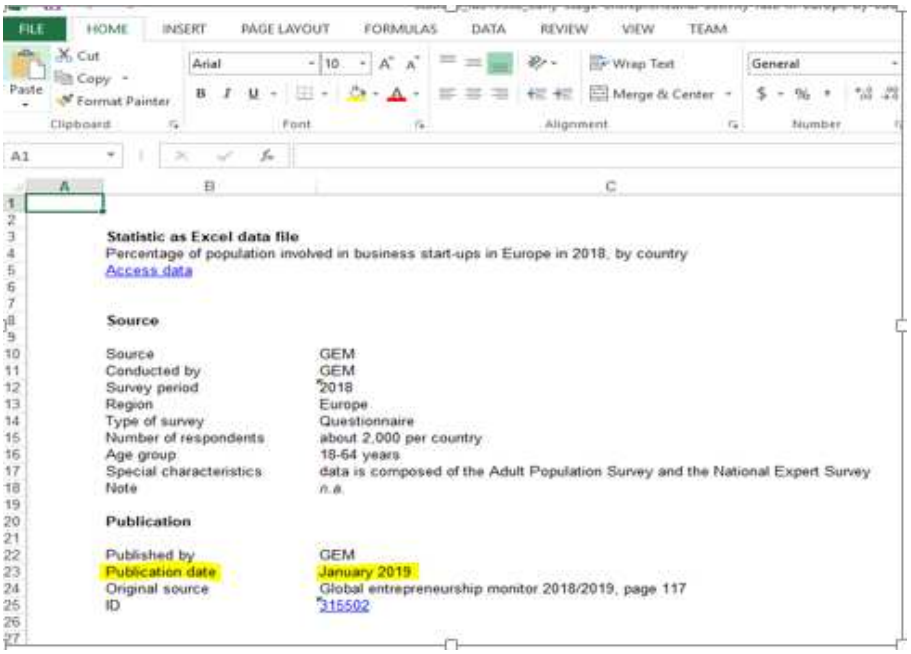


Figure 1: Data Source 1

2.2 Source 2: Kaggle

This dataset gives information about the companies that are in the INC 5000 Europe list.INC 5000 Europe is a list of the 5000 privately held fastest growing companies in Europe for the year 2018.It contains the information about the company like companys name,, url, growth, rank, city, country, industry, Revenue and years on list for all the companies(list has 4900 comapnies).After cleaning the data we get 4988 rows of data((excluding column name)) with 5 columns for Country, City ,Growth Rate ,Company, Revenue in Millions and Industry.Link-`https://www.kaggle.com/msrhossain/2018-inc-5000-europe/activity`

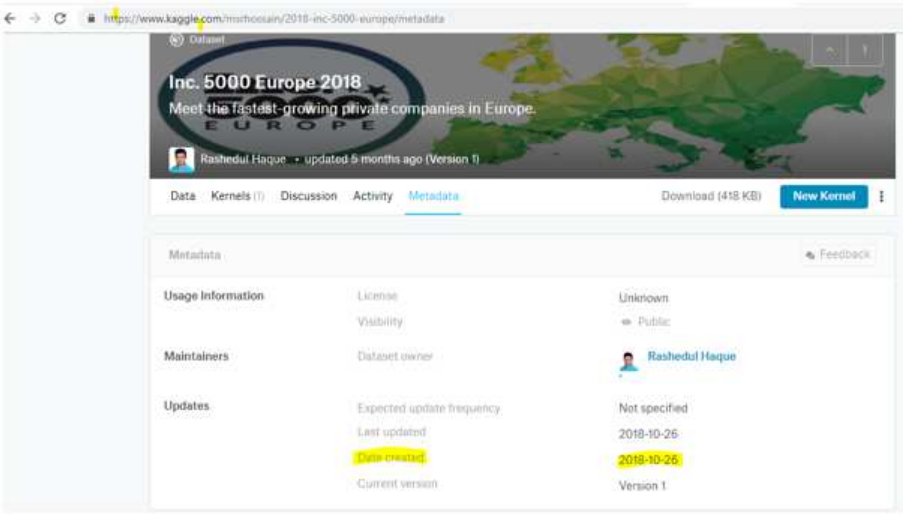


Figure 2: Data Source 2

2.3 Source 3: OECD

This site gives the information about the different types of corporate tax in different European Countries for the year 2018.After applying filters on the data in OECD site and cleaning the data we get 91 columns of data (excluding column name) with four columns for Country, CORP-TAX, Corporate Income tax rate and Tax Rate.Link- `https://stats.oecd.org/index.aspx?DataSetCode=TABLE_II1`

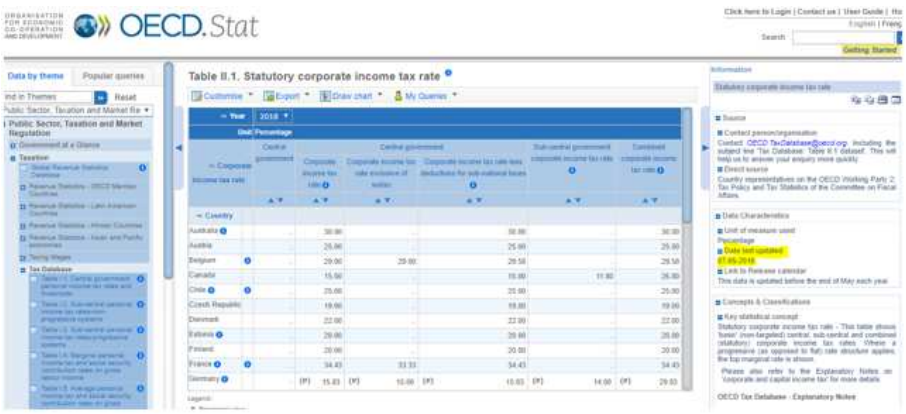


Figure 3: Data Source 3

2.4 Source 4: Eurostat

This site gives data about the unemployment rate in different European Countries by gender (Males ,Females and Total) for the year 2018.After cleaning up the data we get 81 rows of data (excluding column name) and 3 columns for Country .Gender and Unemployment RateLink- `http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=une_rt_a&lang=en`

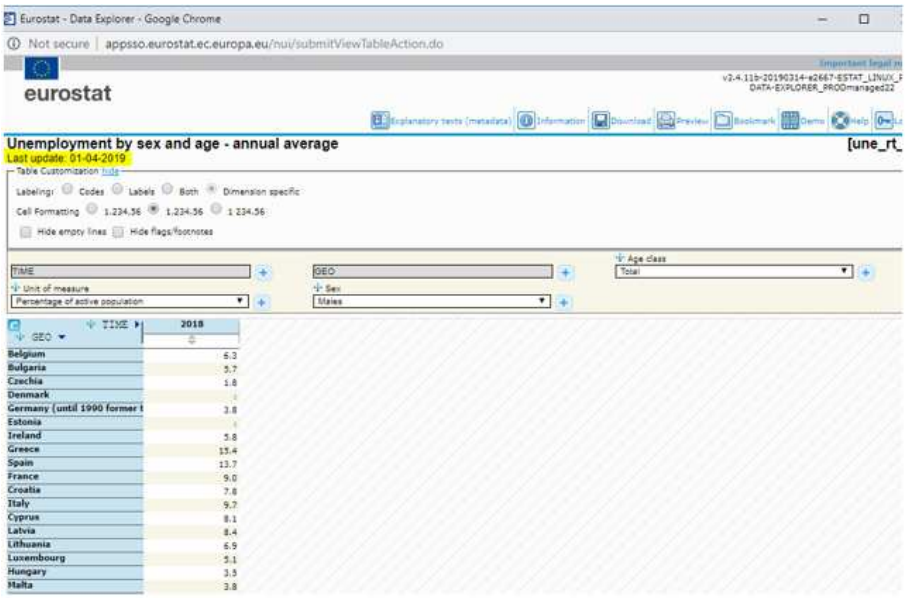


Figure 4: Data Source 4

2.5 Source 5: EU Startup-Monitor Report 2018

This report gives data (page-30) about the percentage of startups that are in seed stage, startup stage, growth stage, later stage and steady stage in different European Countries. None column points out to the percentage of startups that are not in any of the stages and no output column points out to the percentage of startups that are not giving any output. After cleaning up the data we get 22 rows of data (excluding column name) and 7 columns for Country, Seed stage, Startup stage, Growth stage, Later stage, Steady stage and NONE.Link- `http://startupmonitor.eu/EU-Startup-Monitor-2018-Report-WEB.pdf`

Figure 5: Data Source 5

ANNEX FIVE: *Development Stages of Startups*

COUNTRY	SEED STAGE	STARTUP STAGE	GROWTH STAGE	LATER STAGE	STEADY STAGE	NONE	NO INPUT
Austria*	14.2%	44.3%	33.6%	11.0%	2.2%	4.6%	NA
Belgium	3.7%	51.9%	37.0%	0.0%	1.2%	4.9%	1.2%
Czech Republic	11.1%	48.1%	37.0%	3.7%	0.0%	0.0%	0.0%
Denmark	10.6%	36.2%	51.1%	0.0%	0.0%	0.0%	2.1%
France	8.6%	45.7%	37.1%	1.4%	5.7%	1.4%	0.0%
Germany	15.6%	39.5%	39.0%	1.5%	2.9%	1.5%	0.0%
Greece	24.4%	53.7%	17.1%	0.0%	4.9%	0.0%	0.0%
Hungary	39.5%	28.9%	23.7%	0.0%	2.6%	2.6%	2.6%
Ireland	11.3%	46.0%	33.9%	4.8%	2.4%	0.8%	0.8%
Italy	13.9%	55.8%	23.0%	2.4%	3.0%	0.0%	1.8%
Netherlands	9.7%	45.8%	37.5%	1.4%	5.6%	0.0%	0.0%
Poland	16.9%	53.5%	23.9%	0.0%	4.2%	1.4%	0.0%
Portugal	16.7%	51.3%	25.6%	0.0%	5.1%	1.3%	0.0%
Slovakia	14.7%	23.5%	50.0%	2.9%	5.9%	2.9%	0.0%
Slovenia	19.4%	30.6%	38.9%	2.8%	5.6%	2.8%	0.0%
Spain	10.4%	56.3%	24.0%	1.0%	4.2%	3.1%	1.0%
Sweden	7.8%	45.1%	45.1%	0.0%	0.0%	0.0%	2.0%
United Kingdom	8.7%	37.7%	47.8%	1.4%	4.3%	0.0%	0.0%
EU average	**13.2%**	**46.1%**	**33.7%**	**1.5%**	**3.3%**	**1.5%**	**0.7%**
Israel	19.6%	53.6%	19.6%	0.0%	0.0%	5.4%	1.8%
Switzerland	13.5%	50.0%	31.8%	0.7%	0.7%	2.0%	1.4%
Turkey	31.0%	40.5%	19.0%	2.4%	2.4%	2.4%	2.4%

Figure 6: Data Source 5

2.6 Source 6: Wikipedia

Web data extraction is done using R to get Nominal GDP of different European Countries for the year 2018.After extracting and cleaning the data we get 40 rows of data (excluding column name) and 2 columns for Country and GDP2018.Link-`https://en.wikipedia.org/wiki/List_of_sovereign_states_in_Europe_by_GDP_(nominal)`

List of nominal GDP for European countries in billion USD [edit]

This is a sortable list of all European countries by their gross domestic product in US dollars at market or official government exchange rates (nominal GDP). The data for 2018 estimates

2018 Rank	Country	2011	2012	2013	2014	2015	2016	2017	2018[2]
1st	Germany	3,761.142	3,545.946	3,753.687	3,885.440	3,365.293	3,479.232	3,684.816	4,029.140
2nd	United Kingdom	2,511.108	2,655.458	2,721.489	2,945.394	2,858.482	2,660.687	2,624.529	2,808.899
3rd	France	2,865.304	2,682.900	2,809.390	2,843.673	2,420.163	2,466.472	2,583.560	2,794.696
4th	Italy	2,278.376	2,073.972	2,130.997	2,141.937	1,815.759	1,860.152	1,937.894	2,086.911
5th	Russia	2,031.768	2,170.145	2,230.624	2,030.973	1,326.016	1,283.286	1,527.469	1,576.488
6th	Spain	1,489.381	1,340.689	1,369.690	1,383.537	1,199.715	1,237.766	1,313.951	1,437.047
7th	Netherlands	894.576	829.406	866.951	881.033	750.696	777.548	825.745	909.887
8th	Switzerland	696.528	665.898	685.871	712.050	664.603	668.748	678.575	709.118
9th	Sweden	563.110	543.881	579.526	570.137	492.618	514.460	538.575	554.659
10th	Poland	524.104	496.687	526.031	546.644	474.893	471.216	524.886	549.478
11th	Belgium	528.721	499.129	524.970	534.672	454.687	468.148	494.733	536.055
12th	Austria	429.493	407.601	428.456	437.123	374.124	390.961	416.845	459.401
13th	Norway	498.157	509.705	522.349	500.244	389.482	371.075	396.457	441.439
14th	Ireland	237.990	222.089	232.150	246.438	238.031	304.499	333.994	366.448
15th	Denmark	341.498	322.277	335.878	340.806	294.951	306.900	324.484	354.683
16th	Finland	273.925	256.849	268.281	271.155	229.671	238.776	253.244	276.553
17th	Czech Republic	227.307	206.751	208.796	205.658	181.856	195.306	213.189	244.540
18th	Romania	186.113	172.041	191.598	199.950	177.315	187.807	211.315	239.440
19th	Portugal	245.120	216.488	224.983	230.012	199.077	205.269	218.064	237.962
20th	Greece	289.068	249.663	242.306	238.025	195.330	199.720	200.690	218.047

Figure 7: Data Source 6

3 Related Work

There has been notable work on analyzing factors affecting entrepreneurship in the past. One study (Audretsc, Belitski and Desai,2018) analyze the relationship between NBR(National Business Regulations) and City level Entrepreneurship in Europe using a theoretical framework. The results of this study shows that NBR related to entry is partially responsible for growth in city level entrepreneurship. NBR related to property registration discourages entrepreneurship. NBR related to tax has negative relationship with entrepreneurship , as increase in tax payments reduces entrepreneurship. It is found that a city is likely to have 0 entrepreneurship if the no. of tax payments exceeds 35.NBR related to enforcing contracts encourage city level entrepreneurship.

Theres a study(Mrkaji, Murtinu and Scalera,2017) that aims to find out if green startups(companies adopting environment friendly approach or lessening the damage done to the environment) are likely to get venture capital funding.The results show that green ventures will not likely get the venture capital funding compared to the other hi-tech ventures as the funding is limited and it takes a lot of time for green ventures to start earning profit.Green ventures also need extended funding time compared to other hi-tech ventures that are also competing for funding.

Culture in Europe has been identified as a strong influence in the three phases of entrepreneurship that are: beginning stage of business, development stage of a business and stabilization of the business(Gerosa and Tirapani, 2013).

A study (Roper and Scott, 2009) shows that women are more likely to face financial barriers for starting a startup than male based on the data taken from the UK. Another study (Davidaviien, 2016)shows that although migrant setup business at a higher rate than the natives they tend to face a lot more challenges compared to the native people. The result of the study by Davidaviien shows that countries like Netherland, France, Finland and UK based on their EU government e-services are better for migrants to start a business. EU Start-up monitor report 2018(Steigertahl, Mauer and Say,2018) gives us the information about start-ups all across the Europe, the data is generated in three fields which are Profile, Internationalization, and Cooperation. From the research it is found that the average start-up founder is usually a male (82.8 percent of the times),

he has a university degree (84.8 percent of the times) and at present is 38 years old(35 years old at the time of start-up business).The study shows that the biggest motivation for finding a start-up was self-fulfillment(79.1 percent of the times) while Independence was second at 62.8 percent.The report mentions that on average 18,015 jobs would be created in Europe in the next 12 months based on the samples they have collected. The reports have identified that 88 percent of start-ups want to expand to other countries (internationalization) in the next 12 months so that they can earn profit and there can be a continuous cash flow for the business. Another factor, cooperation focuses on making both Big corporate companies and start-up work together. The study has found that there are already 71.1 start-ups are in cooperation with big companies.

The first dataset from Statista is taken from the Global Entrepreneurship Monitor(GEM) report for the year 2018.This report gives the entrepreneurship information for all the countries based on the survey and studies conducted by GEM. My corporate tax data is from OECD (Organisation for Economic Co-operation and Development),this data is used by the organization for making policies to improve the economic and social well-being of people in the world.My unemployment datset is from Eurostat which provides statistical data to Europe Union so that decisions can be taken by the EU based on the data provided. My dataset from EU Startup monitor 2018 is the result of the research and survey conducted to get the figures of companies in different stages of start-ups in different country. The datset with GDP information is used to compare the economic growth between the countries and analyze the reason for having a lower or higher GDP.

By undertaking this project I wish to find out the correlation between entrepreneurship, corporate tax and unemployment in European countries, I also want to find if the countries with more fastest growing start-ups(INC 5000 Europe companies) and having more no. of start-ups in growth stage have a correlation with corporate tax of the country. Lastly I want to see if having higher number of fastest growing start-ups affects the economy of the country(GDP).So my model doesnt just analyze the factors (corporate tax) that affect the entrepreneurship rate of the countries it also takes into account the factors that gets affected by the entrepreneurship in a country(unemployment and GDP).

4 Data Model

The star schema is the simplest form of schema in the Data warehouse. It is known as star schema because it forms a star like shape with Fact table and dimensions table. Fact table lies in the middle while dimensions table surrounds the fact table exactly like end points of a star.(DW4U) Fact table- A fact table consists of foreign keys and numerical facts. Dimension table-A Dimension table consists of a primary key and attributes of data which are usually text. A primary key of dimension table is linked with the foreign key of the fact table, making every table interconnected with each other.

For my Data warehouse project I had made a star schema using one fact table and four dimension table.

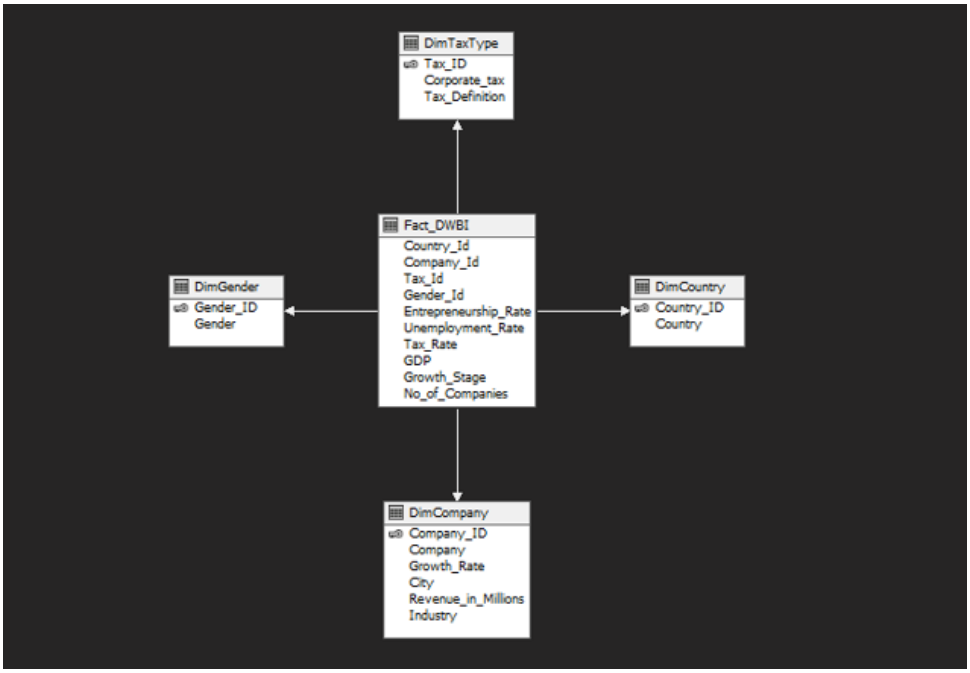


Figure 8: Star Schema

DimTaxType-This dimension consists of a primary key Tax-ID,Corporate-tax that contains abbreviation of tax type and Tax-Definition that properly defines the Corporate-tax abbreviations.The id is auto generated (from SQL query) for every entry in the table by an increment of 1.There are 5 types of Corprate-tax in the dimension. This dimension is created using the dataset from OECD about corporate taxes of different countries.

Results Messages

	Tax_ID	Corporate_tax	Tax_Definition
1	1	CIT_RATE	Corporate income tax rate
2	2	CIT_RATE_EXCL_SURTAX	Corporate income tax rate exclusive of surtax
3	3	CIT_RATE_LESS_SUB_NAT	Corporate income tax rate less deductions for su...
4	4	COMB_CIT_RATE	Combined corporate income tax rate
5	5	SUB_CENT_GOV_CIT_RATE	Sub-central government corporate income tax rate

Figure 9: Dimension DimTaxType

DimCountry-This dimension consists of a primary key Country-ID and Country(Countrys name).The id is auto generated (from SQL query) for every entry in the table by an increment of 1.This dimension is created from the GDP dataset that is received after web scraping of Wikipedia page containing GDP information. This dimension has 40 rows of data.

Results Messages

	Country_ID	Country
1	1	Albania
2	2	Austria
3	3	Belarus
4	4	Belgium
5	5	Bosnia and Herzegovina
6	6	Bulgaria
7	7	Croatia
8	8	Czech Republic
9	9	Denmark
10	10	Estonia
11	11	Finland

Figure 10: Dimension DimCountry

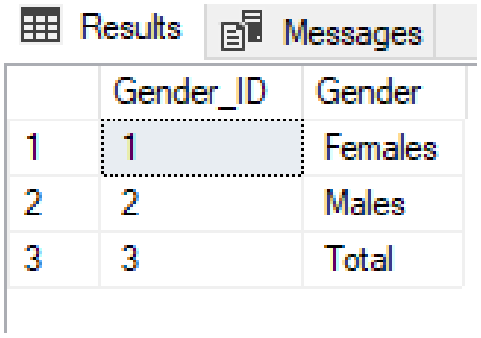
DimCompany-This dimension consists of a primary key Company-ID, Company, Growth-Rate ,City , Revenue-in-Millions and Industry(type of industry a company is like IT industry,Travel Hospitality,Construction etc). The id is auto generated (from SQL query) for every entry in the table by an increment of 1.This dataset is created from INC5000 Europe dataset that was taken from Kaggle.This dimension has a data of 4988 rows.

Results Messages

	Company_ID	Company	Growth_Rate	City	Revenue_in_Millions	Industry
4...	4978	Olgerdin Egill Skallagrimsson hf.	54.6539	Reykjavik	134590931	Food & Beverage
4...	4979	BIPSO	54.4752	Singen	73410483	Health
4...	4980	Mach Media NV	53.8158	Gent	2383610	Advertising & Marketing
4...	4981	Business Support Spa	53.5183	milano	2181807	Business Products & Services
4...	4982	S.C. SUNIMPROF ROTTAPRINT ...	52.2225	Cluj-Napoca	16436475	Business Products & Services
4...	4983	Marketshot	51.7862	Paris	7010974	Advertising & Marketing
4...	4984	Accso - Accelerated Solutions Gm...	51.3349	Darmstadt	8457839	IT Services
4...	4985	ROMFRACHT SRL	49.1209	Ilfov	31017872	Logistics & Transportation
4...	4986	BAA Training	48.7292	Vilnius	11177000	Business Products & Services
4...	4987	interworks.cloud	48.2759	Thessaloniki	2150000	Software
4...	4988	D.R.I.M. Daniel SRL	47.9696	Pitesti	59176000	Logistics & Transportation

Figure 11: Dimension DimCompany

DimGender-This dimension consists of a primary key Gender-ID and Gender.The id is autogenerated (from SQL query) for every entry in the table by an increment of 1.This dimension is made from the unemployment(by gender) dataset that was taken from Eurostat.This dimension has 3 rows of data-Males,Female and Total(both male and female).

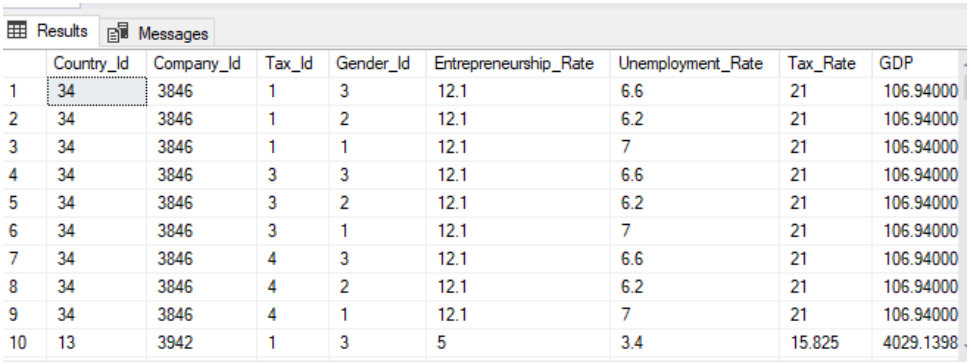


Results Messages

	Gender_ID	Gender
1	1	Females
2	2	Males
3	3	Total

Figure 12: Dimension DimGender

Fact-DWBI- The fact tables contains Country-Id, Company-Id, Tax-Id, Gender-Id, Entrepreneurship-Rate, Unemployment-Rate, Tax-Rate, GDP, Growth-Stage and No-of-Companies. Country-Id is a foreign key that has a relation with Primary Key Country-ID of dimension DimCountry.Company-Id is a foreign key that has a relation with Primary Key Company-ID of dimension DimCompany.Tax-Id is a foreign key that has a relation with Primary Key Tax-ID of dimension DimTaxType.Gender-Id is a foreign key that has a relation with Primary Key Gender-ID of dimension DimGender.This primary key and foreign key relationship between fact table and dimension tables interconnects all the tables together. Entrepreneurship-Rate, Unemployment-Rate, Tax-Rate, GDP, Growth-Stage and No-of-Companies are the numerical facts related to the country.Entrepreneurship rate gives the percentage of early stage entrepreneurship activity in different countries.Unemployment-Rate gives the percentage of unemployed people in different countries.Tax-Rate is the rate of tax (in percentage) in different countries.GDP gives the Nominal GDP of each country in Million US Dollar.Growth-Stage is the percentage of start-up companies in a country that are in growth stage. No-of-Companies gives the total count of the companies that are in INC 5000 Europe list for each country.



Results Messages

	Country_Id	Company_Id	Tax_Id	Gender_Id	Entrepreneurship_Rate	Unemployment_Rate	Tax_Rate	GDP
1	34	3846	1	3	12.1	6.6	21	106.94000
2	34	3846	1	2	12.1	6.2	21	106.94000
3	34	3846	1	1	12.1	7	21	106.94000
4	34	3846	3	3	12.1	6.6	21	106.94000
5	34	3846	3	2	12.1	6.2	21	106.94000
6	34	3846	3	1	12.1	7	21	106.94000
7	34	3846	4	3	12.1	6.6	21	106.94000
8	34	3846	4	2	12.1	6.2	21	106.94000
9	34	3846	4	1	12.1	7	21	106.94000
10	13	3942	1	3	5	3.4	15.825	4029.1398

Figure 13: Fact table FactDWBI

5 Logical Data Map

This section explains the source of all the columns that we are using in our star schema and mention any transformation that we might have done on it.

Table 2: Logical Data Map describing all transformations, sources and destinations for all components of the data model illustrated in 8:Star Schema

Source	Column	Destination	Column	Type	Transformation
6	Country	Dim-Country	Country	Dimension	This column had to be cleaned using R to remove the character and change the name of some Countries from Slovakia to Slovak Republic, UnitedKingdom to United Kingdom and Switzerland had spaces before it which also had to be removed.
6	Country-ID	Dim-Country	Country-ID	Dimension	This column was made to have numeric values that auto increment by 1 based on the entries of column Country.
2	Company-ID	Dim-Company	Company-ID	Dimension	This column was made to have numeric values that auto increment by 1 based on the row entries in the table.
2	Company	Dim-Company	Company	Dimension	Only changed the column name from company to Company for the raw data.
2	Growth Rate	Dim-Company	Growth-Rate	Dimension	This column was converted to double using R as it was a character column earlier.
2	city	Dim-Company	City	Dimension	Only changed the column name from city to City for the raw data.
2	Revenue in Millions	Dim-Company	Revenue-in-Millions	Dimension	This column was converted to double using R as it was a character column earlier.
2	industry	Dim-Company	Industry	Dimension	Only changed the column name from industry to Industry for the raw data.

Continued on next page

Table 2 – *Continued from previous page*

Source	**Column**	**Destination**	**Column**	**Type**	**Transformation**
4	Gender-ID	DimGender	Gender-ID	Dimension	This column was made to have numeric values that auto increment by 1 based on the entries of column Gender.
4	Gender	DimGender	Gender	Dimension	As this column had repetitive values for each country the Sql query was used to insert only distinct values.
3	Tax Type	DimTaxType	Corporate-tax	Dimension	No transformation was done on this column.
3	Tax Definition	DimTaxType	Tax-Definition	Dimension	No transformation was done on this column.
3	Tax-ID	DimTaxType	Tax-ID	Dimension	This column was made to have numeric values that auto increment by 1 based on the row entries in the table.
1	Entreprenu-rship Rate	Fact-DWBI	Entrepreneurship-Rate	Fact	This column was converted to double using R as it was a character column earlier.
4	Unemploy-mentRate	Fact-DWBI	Unemployment-Rate	Fact	This column was converted to double using R as it was a character column earlier.
3	Tax Rate	Fact-DWBI	Tax-Rate	Fact	This column was converted to double using R as it was a character column earlier.
5	Growth Stage	Fact-DWBI	Growth-Stage	Fact	This column was converted to double using R as it was a character column earlier.
2	Companies	Fact-DWBI	No-of-Companies	Fact	A new table is created from source 2 that counts the number of companies in the list for each country.
6	GDP2018	Fact-DWBI	GDP	Fact	This column was converted to double using R as it was a character column earlier.

6 ETL Process

ETL is the abbreviation of Extract, Load and Transform. It is used to construct a Data warehouse by pulling a data from the original source and placing them in the targeted database (data warehouse). The three functions of ETL are explained below: Extract - In this the data is extracted (or read) from the source. Transform - The extracted data is then converted into an appropriate format (transformation) for analyzing. Load - The transformed data is then stored (load) in the data warehouse.

I have 6 Execute Process for the ETL Process of my project. All the Execute Process are described below-

1.GDP ETL- In this ETL process I have provided the R code (gdpwiki.R) in the SSIS Execute Process Task. I have extracted the data (using htmltab function of R) and cleaned the data (removed unnecessary characters, changed the data type of the column and fixed the names of columns and countries) so that it could become suitable for analyzing for my BI Queries. Then using Flat File Source I have uploaded the transformed csv file (EUgdp.csv) in SSIS and using OLE DB Destination that file is loaded into the data warehouse of my project.

2.Corporate tax data ETL- In this ETL process I have provided the R code (corptax.R) in the SSIS Execute Process Task. The R code read the csv file that I had downloaded from Eurostat and cleaned the data (removed unnecessary columns, changed the data type of column and fixed the names of columns) so that it could become suitable for analyzing for my BI Queries. Then using Flat File Source I have uploaded the transformed csv file (corporatetax.csv) in SSIS and using OLE DB Destination that file is loaded into the data warehouse of my project.

3. Unemployment ETL- In this ETL process I have provided the R code (unemplymentinfo.R) in the SSIS Execute Process Task. The R code read the csv file that I had downloaded from OECD and cleaned the data (removed unnecessary columns, changed the data type of column and fixed the names of columns) so that it could become suitable for analyzing for my BI Queries. Then using Flat File Source I have uploaded the transformed csv file (unemployed.csv) in SSIS and using OLE DB Destination that file is loaded into the data warehouse of my project.

4.INC5000 Company ETL- In this ETL process I have provided the R code (companyinfo.R) in the SSIS Execute Process Task. The R code read the excel file that I had downloaded from Kaggle and cleaned the data (removed unnecessary columns, changed the data type of columns and fixed the names of columns) so that it could become suitable for analyzing for my BI Queries. Then using Flat File Source I have uploaded the transformed csv file (eucompany.csv) in SSIS and using OLE DB Destination that file is loaded into the data warehouse of my project.

5.Entrepreneurship rate ETL- In this ETL process I have provided the R code (entreinfo.R) in the SSIS Execute Process Task. The R code read the excel file that I had downloaded from Statista and cleaned the data (changed the data type of column and fixed the names of column) so that it could become suitable for analyzing for my BI Queries. Then using Flat File Source I have uploaded the transformed csv file (euentrepreneurship.csv) in SSIS and using OLE DB Destination that file is loaded into the data warehouse of my project.

6. Entrepreneurship stages ETL- In this ETL process I have provided the R code (newscript.R) in the SSIS Execute Process Task. The R code read the pdf file that I had downloaded from EU and cleaned the data (removed unnecessary characters, changed the

data type of columns and fixed the names of column) so that it is suitable for analyzing for my BI Queries. Then using Flat File Source I have uploaded the transformed csv file (startupstage.csv) in SSIS and using OLE DB Destination that file is loaded into the data warehouse of my project.

7 Application

The cube that was deployed is used in Tableau for the visualization of the 3 BI Queries below:

7.1 BI Query 1: : Does the country with higher corporate tax has lower entrepreneurship rate and higher female unemployment than the country with lower corporate tax?

This is a non trivial and multi-dimensional query.Three data set are being used in this BI Query- first data source is from statista about entrepreneurship rate in different countries , second data set is from OECD about corporate tax of different countries and third data set is from Eurostat about unemployment rate in different Countries. The dimensions used are DimGender, DimCountry and DimTaxType(to get female unemployment, combined corporate tax rate and Country name). France has corporate tax of 34.43 percent, female unemployment rate of 9.10 percent and entrepreneurship rate of 6.10 percent. Ireland has corporate tax of 12.5 percent,female unemployment rate of 5.70 percent and entrepreneurship rate of 9.60 percent. We can observe similar cases for other countries(Greece,Slovenia,United Kingdom) as well that as the corporate tax increases the entrepreneurship rate is increased and female unemployment is decreased. So from our first BI query we know that the country that has lower corporate tax promotes entrepreneurship which in turn lowers the unemployment rate of females in the country. Figure 14.

Figure 14: Results for BI Query 1

7.2 BI Query 2: Do the countries with lower corporate tax have more number of companies in INC 5000 Europe list and more percentage of start-ups in growth rate?

This is a non trivial and multi-dimensional query.Three data sets are being used in this BI Query- first data set is from EU Startup- Monitor Report 2018 about the percentage

of companies in different start-up stages for each country, second data set is from OECD about corporate tax of different European countries and third data source is from Kaggle about companies in INC5000 Europe list.The dimensions used are DimCountry and DimTaxType(to get combined corporate tax rate and Country name).

After carefully observing the visualization we can see that the countries with higher corporate tax have more no. of private fastest growing companies and more percentage of start-ups in the growth stage then the companies with lower corporate tax. Greece and France that have high corporate taxes have more no. of companies in INC 5000 list and more no. of start-ups in growth stage then Ireland,Hungary and Poland that have low corporate taxes.This case could be seen for other countries as well. So the answer for the second Bi query is that the countries with lower corporate tax have do not have more number of companies in INC 5000 Europe list and more percentage of start-ups in growth rate

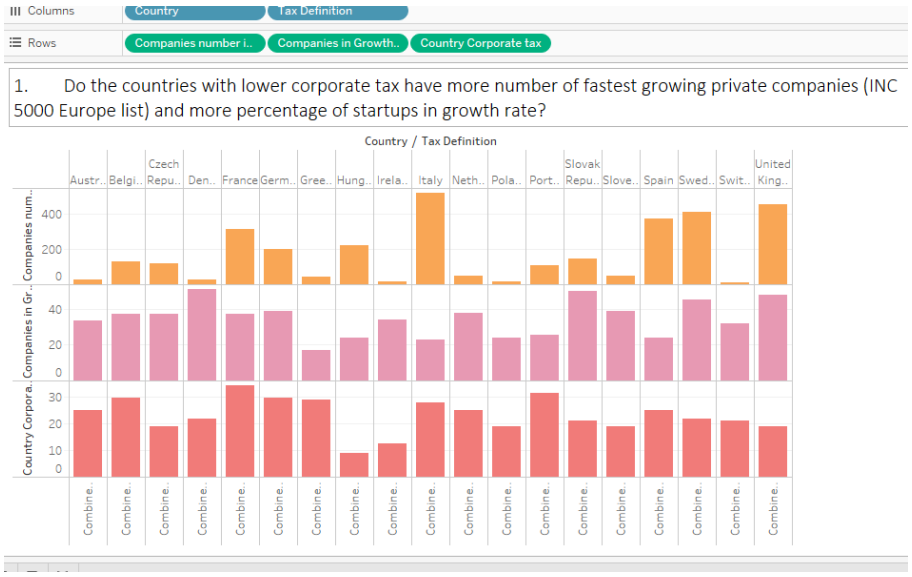


Figure 15: Results for BI Query 2

7.3 BI Query 3: Do the countries that have more number of companies in the INC 5000 Europe list have higher GDP ?

This a non trivial query and two data sets are used in this BI Query.First dataset is from Kaggle about companies in INC5000 Europe list and second dataset is from Wikipedia about GDP of European countries.

From the visualization we can observe that as GDP decrease the no. of companies for a country also decreases.So we can say that the fastest growing companies in a country are a contributing factor to the country's GDP and increase in these companies would increase the GDP of a country.

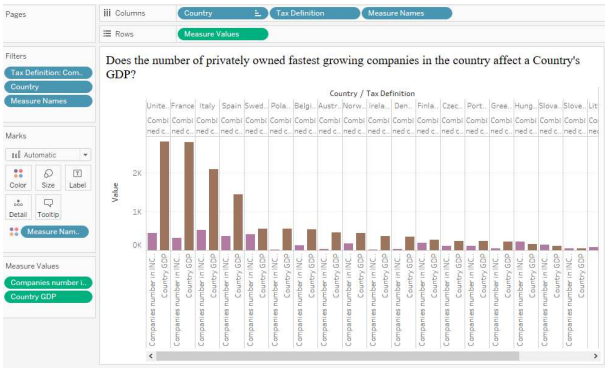


Figure 16: Results for BI Query 3

7.4 Discussion

The first BI Query finds that country with higher corporate tax (France) has lower entrepreneurship rate and higher female unemployment rate while the country with lower corporate tax (Ireland) has higher entrepreneurship rate and lower female unemployment rate.This is also the case for the country with the second highest corporate tax(Greece) and for the country with second lowest corporate tax(United Kingdom).Thus it can be assumed from the results that lower corporate in a country encourages entrepreneurship and lowers the unemployment rate. From the second BI Query we can see a reverse relation between corporate tax, no. of companies in INC 5000 and percentage of start-ups in Growth stage for each country .We know from our second BI query that country having higher corporate tax has more no. of privately owned companies and more no. of start-ups in growth rate.From the third BI Query we know that the countries that have higher number of companies in INC 5000 Europe list have higher GDP. The Start-ups play a vital role in the economic growth of the nation and from the third BI Query we can see that there is a positive relationship between GDP and the no. of fastest growing private companies (INC 5000 Europe).

8 Conclusion and Future Work

The first Bi query concluded that countries with higher corporate tax have lower entrepreneurship rate and higher female unemployment rate.The second BI Query did not find any correlation among corporate tax, no. of companies in INC 5000 and percentage of start-ups in Growth stage.the third Bi Query found a correlation(positive relationship) between GDP and number of companies in INC 5000 Europe list.

For future Corruption Perception Index of different countries can be used to find the correlation with entrepreneurship rate or number of start-ups generated in a country.From this we can see if corruption affects entrepreneurship rate of countries.Private Equity Funding can also be used to find if countries having larger Private Equity Funding have any correlation with the no. of start-ups generated in that country. The data for European Unicorn companies can also be taken to see which country in Europe have more unicorns.

9 References

Audretsch, D. B., Belitski, M. Desai, S. (2018), 'National business regulations and cityentrepreneurship in europe: A multilevel nested analysis',Entrepreneurship Theory and Practice.

Gerosa, A. Tirapani, A. N. (2013), 'The culture of entrepreneurship: Creating your own job',European View12(2), 205214.

Mrkajic, B., Murtinu, S. Scalera, V. G. (2017), 'Is green the new gold? venture capitaland green entrepreneurship',Small Business Economics52(4), 929950.

Steigertahl, L. Mauer, R. (2018), 'EU startup monitor'.

Roper, S. Scott, J. M. (2009), 'Perceived financial barriers and the start-up decision',International Small Business Journal: Researching Entrepreneurship21(2), 149171.

Vida, D. Ingrida, L. (2016), 'Migrant entrepreneurship in europe: Challenges andopportunities'.

Cobizmag.com. (2019). The role of startups in the economy. [online] Available at: http://www.cobizmag.com/Trends/The-role-of-startups-in-the-economy/ [Accessed 12 Apr. 2019].

Smale, T. (2019). Why Europe Is the Next Major Startup Hub. [online] Entrepreneur. Available at: https://www.entrepreneur.com/article/307924 [Accessed 12 Apr. 2019].

Appendix

R code example

```
# R code for entrepreneurship rate (statista)
library(readxl)
statistic_id315502_early_stage_entrepreneurial_activity _rate_in_europe_by_
country_2018<-read_excel("statistic_id315502_early-stage-
entrepreneurial-activity-
rate-in-europe-by-country-2018.xlsx")
euentrepreneurship<-statistic_id315502_early_stage_entrepreneurial_activity_
rate_in_europe_by_country_2018[c(1:2)]
names(euentrepreneurship)[names(euentrepreneurship) =="Early-stage-
entrepreneurial-activity-rate-in-Europe-
by-country-2018"] <- "Country"
names(euentrepreneurship)[names(euentrepreneurship) =="...2"] <-
"Entreprenurship␣Rate"
colSums(is.na(euentrepreneurship))
euentrepreneurship<-na.omit(euentrepreneurship)
head(euentrepreneurship)
names(euentrepreneurship)
euentrepreneurship$`Entreprenurship Rate` <-as.double(
as.character(euentrepreneurship$`Entreprenurship Rate`))
write.csv(euentrepreneurship,file="euentrepreneurship.csv",row.names=FALSE)

# R code for unemployment (Eurostat)
library(readr)
une_rt_a_1_Data <- read_csv("une_rt_a_1_Data.csv")
```

```
unemployed<-une_rt_a_1_Data[c(2,5,6)]
head(unemployed)
colSums(is.na(unemployed))
unemployed <- lapply(unemployed, gsub, pattern="until-1990-
former-territory-of-the-FRG", replacement='')
unemployed <- lapply(unemployed, gsub, pattern="\\(\\)", replacement='')
unemployed <- lapply(unemployed, gsub,pattern="Germany_",
replacement="Germany")
names(unemployed)[names(unemployed) == "GEO"] <-"Country"
names(unemployed)[names(unemployed) == "SEX"] <- "Gender"
names(unemployed)[names(unemployed) == "Value"] <- "UnemploymentRate"
unemployed <- lapply(unemployed, gsub, pattern="Slovakia",
replacement="Slovak-Republic")
unemployed$UnemploymentRate <- as.double(as.character(
unemployed$UnemploymentRate))
names(unemployed)
str(unemployed)
write.csv(unemployed,file="unemployed.csv",row.names=FALSE)


#R code for INC 5000 Europe list (Kaggle)
library(readxl)
inc500euSummary <- read_excel("inc500euSummary.xlsx")
company<- inc500euSummary[c(6,3:5,7:8)]
sum(is.na(company))
colSums(is.na(company))
company<-na.omit(company)
names(company)[names(company) == "country"] <- "Country"
names(company)[names(company) == "city"] <- "City"
names(company)[names(company) == "growth"] <- "Growth-Rate"
names(company)[names(company) == "company"] <- "Company"
names(company)[names(company) == "revenue"] <-"Revenue-in-Millions"
company <- lapply(company, gsub, pattern="Slovakia",
replacement="Slovak-Republic")
names(company)[names(company) == "industry"] <- "Industry"
company$'Growth-Rate' <- as.double(as.character(company$'Growth-Rate'))
company$'Revenue-in-Millions' <-as.double(as.character(company$'Revenue-
in-Millions'))
write.csv(company,file="eucompany.csv",row.names=FALSE)

#R code for corporate tax (OECD)
library(readr)
TABLE_II1_17032019130716872 <- read_csv("TABLE_II1_17032019130716872.csv")
my.data<-TABLE_II1_17032019130716872[c(2,3,4,13)]
head(my.data)
names(my.data)[names(my.data) == "Value"] <- "Tax-Rate"
names(my.data)[names(my.data) == "CORP_TAX"] <- "Tax-Type"
names(my.data)[names(my.data) == "Corporate-income-tax-rate"] <-
"Tax-Definition"
my.data$'Tax-Rate' <- as.double(as.character(my.data$'Tax-Rate'))
colSums(is.na(my.data))
```

```
str(my.data)
write.csv(my.data,file="corporatetax.csv",row.names=FALSE)

#R code for GDP (wikipedia)
install.packages("htmltab", repos = "http://cran.us.r-project.org")
library(XML)
library(stringr)
library(dplyr)
library(htmltab)
library(readr)
url = "https://en.wikipedia.org/wiki/List_of_sovereign_states_in_Europe
_by_GDP_(nominal)"
countries <- htmltab(doc=url, which=2)
my.data.country<-countries
names(my.data.country)
finaldata<-my.data.country[c(2,10)]
names(finaldata)[names(finaldata) == "Country "] <- "Country"
names(finaldata)[names(finaldata) == "2018"] <- "GDP2018"
names(finaldata)
finaldata <- lapply(finaldata, gsub, pattern='  ', replacement='')
finaldata <- lapply(finaldata, gsub, pattern=',', replacement='')
finaldata <- lapply(finaldata, gsub, pattern="UnitedKingdom",
replacement="United-Kingdom")
finaldata <- lapply(finaldata, gsub, pattern="Slovakia",
replacement="Slovak-Republic")
finaldata <- lapply(finaldata, gsub, pattern="Russia",
replacement="Russian-Federation")
finaldata <- lapply(finaldata, gsub, pattern="[[:space:]]Switzerland",
replacement="Switzerland")
finaldata$Country <-substr(finaldata$Country, 2,
nchar(finaldata$Country))
finaldata$GDP2018 <- as.double(as.character
(finaldata$GDP2018))
head(finaldata)
head(D)

#R code for growth stage(EU monitor 2018 report table pdf extraction)
install.packages("rJava", repos = "http://cran.us.r-project.org")
install.packages("tabulizer", repos = "http://cran.us.r-project.org")
install.packages("data.table", repos = "http://cran.us.r-project.org")
install.packages("supernova", repos = "http://cran.us.r-project.org")
install.packages("miscTools", repos = "http://cran.us.r-project.org")
Sys.setenv(JAVA_HOME='C:\\Program␣Files\\Java\\jdk1.8.0_201')
library(rJava)
library(tabulizer)
library(data.table)
library(supernova)
library(miscTools)
library(dplyr)
location <- 'C:\\Users\\Admin\\Documents\\EU-Startup-Monitor-2018-Report
-WEB.pdf'
```

```
out <- extract_tables(location,pages = 32,output = "data.frame",
header = TRUE)
head(out)
out$Country<- c("Belgium", "Czech-Republic","Denmark","France","Germany",
"Greece","Hungary","Ireland","Italy","Netherlands","Poland","Portugal",
"Slovak-Republic","Slovenia","Spain","Sweden","United-Kingdom","EU-average"
,"NA","Israel","Switzerland")
str(out)
df<-data.frame(out)
names(df)
df <- add_row(df, X14.2. = "14.2%", X44.3. = "44.3%", X33.6.= "33.6%",
X11.0. = "11.0%", X2.2. = "2.2%", X4.6.= "4.6%", Country= "Austria")
df <- add_row(df, X14.2. = "31.0%", X44.3. = "40.5%", X33.6.= "19.0%",
X11.0. = "2.4%", X2.2. = "2.4%", X4.6.= "2.4%", Country= "Turkey")
str(df)
names(df)[names(df) == "X14.2."] <- "Seed-Stage"
names(df)[names(df) == "X44.3."] <- "Startup-stage"
names(df)[names(df) == "X33.6."] <- "Growth-Stage"
names(df)[names(df) == "X11.0."] <- "Later-Stage"
names(df)[names(df) == "X2.2."] <- "Steady-Stage"
names(df)[names(df) == "X4.6."] <- "NONE"
df<-df[c(7,1:6)]
write.csv(df,file="startupstage.csv",row.names=FALSE)
library(readr)
startupstage <- read_csv("startupstage.csv")
df <- startupstage
sum(is.na(df))
colSums(is.na(df))
df<-na.omit(df)
df <- lapply(df, gsub, pattern='%', replacement='')
out <- lapply(out, gsub, pattern="Austria\\*", replacement="Austria")
df$'Seed-Stage' <- as.double(as.character(df$'Seed-Stage'))
df$'Startup-stage' <- as.double(as.character(df$'Startup-stage'))
df$'Growth-Stage' <- as.double(as.character(df$'Growth-Stage'))
df$'Later-Stage' <- as.double(as.character(df$'Later-Stage'))
df$'Steady-Stage' <- as.double(as.character(df$'Steady-Stage'))
df$'NONE' <- as.double(as.character(df$'NONE'))
names(df)
str(df)
write.csv(df,file="startupstage.csv",row.names=FALSE)
write.csv(finaldata,file="EUgdp.csv",row.names=FALSE)

#SQL Queries for Dimension Table creation

DROP TABLE IF EXISTS [DimCountry] CREATE TABLE
[DimCountry]([Country_ID] int NOT NULL IDENTITY (1, 1)
PRIMARY KEY, [Country] varchar(50));

DROP TABLE IF EXISTS [DimGender] CREATE TABLE
[DimGender]([Gender_ID] intNOT NULL IDENTITY (1, 1)
PRIMARY KEY, [Gender] varchar(50));
```

```
DROP TABLE IF EXISTS [DimTaxType] CREATE TABLE
 [DimTaxType]([Tax_ID] int NOT NULL IDENTITY (1, 1)
 PRIMARY KEY,[Corporate_tax] varchar(50),[Tax_Definition] varchar(max) );

DROP TABLE IF EXISTS [DimCompany] CREATE TABLE [DimCompany]([Company_ID]
int NOT NULL IDENTITY (1, 1) PRIMARY KEY, [Company] varchar(max),
[Growth_Rate] float,[City] varchar(50), [Revenue_in_Millions] float,
[Industry] varchar(max));

#SQL Queries for Inserting in Dimension Table

insert into dbo.DimCountry
([Country])
select DISTINCT r.[Country]from dbo.EUGDP r

insert into dbo.DimGender
([Gender])
select DISTINCT r.[Gender]
from dbo.Unemployment r

insert into dbo.DimTaxType
([Corporate_tax],[Tax_Definition])
select DISTINCT r.[Tax Type]
,r.[Tax Definition]
from dbo.EUtax r

insert into dbo.DimCompany
([Company],
[Growth_Rate],
[City],
[Revenue_in_Millions],
[Industry])
select r.[Company], r.[Growth Rate], r.[City],
 r.[Revenue in Millions], r.[Industry]
from dbo.Comapnies r

#SQL Query  for counting no. of  companies for each country

DROP TABLE IF EXISTS [Countcomapny] CREATE TABLE CountCompany(
   [Country] varchar(50),
   [Number_of_Companies] float

);

INSERT INTO dbo.CountCompany(Country,Number_of_Companies)
 select c.Country,count(c.Company)
from Comapnies c
group by c.Country;
select c.Country,count(c.Company)
from Comapnies c
```

```
group by c.Country;

#SQL Query for creating a fact table

DROP TABLE IF EXISTS [Fact_DWBI]
CREATE TABLE [Fact_DWBI](
Country_Id int Foreign Key References DimCountry(Country_ID),
Company_Id int Foreign Key References DimCompany(Company_ID),
Tax_Id int Foreign Key References DimTaxType(Tax_ID),
Gender_Id int Foreign Key References DimGender(Gender_ID),
[Entrepreneurship_Rate] float,
[Unemployment_Rate] float,
[Tax_Rate] float,
[GDP] float,
[Growth_Stage] float,
[No_of_Companies] float
);



#SQL Query for inserting values in a fact table

INSERT INTO [dbo].[Fact_DWBI]
   (
[Country_Id],
[Company_Id],
[Tax_Id],
[Gender_Id],
[Entrepreneurship_Rate],
[Unemployment_Rate],
[Tax_Rate],
[GDP],
[Growth_Stage],
[No_of_Companies])
select [Country_Id]
,[Company_Id]
,[Tax_Id]
,[Gender_Id]
,c.[Entreprenurship Rate],e.
[UnemploymentRate],d.[Tax Rate],k.[GDP2018],
b.[Growth Stage],x.[Number_of_Companies]
from [dbo].[Comapnies] a LEFT JOIN [dbo].[Comapny Stage]
b on b.[Country]=a.[Country]
LEFT JOIN [dbo].[Entrepreneurship] c on c.[Country]=a.[Country]
LEFT JOIN [dbo].[CountCompany] x on x.[Country]=a.[Country]
LEFT JOIN [dbo].[EUtax] d on d.[Country]=a.[Country]
LEFT JOIN [dbo].[Unemployment] e on e.[Country]=a.[Country]
LEFT JOIN [dbo].[EUGDP] k on k.[Country]=a.[Country]
LEFT JOIN [dbo].[DimCompany] f on f.[Company]=a.[Company]
LEFT JOIN [dbo].[DimCountry] m on m.Country = a.[Country]
LEFT JOIN [dbo].[DimGender] r on r.Gender=e.[Gender]
```

```
LEFT JOIN [dbo].[DimTaxType] o on o.Corporate_tax=d.[Tax Type]
```