Database Project Queries
Sport Product Manufacturing

Pforzheim University

Author Alice Zimpfer

Abstract

Documentation of Queries for the Database Project about the topic

“Sport Product Manufacturing”.

January 2024

Zimpfer

1. Introduction

This document will define my queries in natural language and relational algebra, it will also interprete the query results.

To see my queries in sql visit and open the .sql file:
https://github.com/hochschule-pforzheim/project-wt2324-Arity-jpg/tree/main/queries
To see my model visit and open the picture:
https://github.com/hochschule-pforzheim/project-wt2324-Arity-jpg/tree/main/model
To see my relational schema visit and read the documentation.md:
https://github.com/hochschule-pforzheim/project-wt2324-Arity-jpg/blob/main/documentation/documentation.md

2. Natural Language and Relational Algebra

1) In this Section information about countries containing factories is explored.

1.1) Using Address_Mondial: Count the Factories per Country. Note, it has an aggregate function.

The result shows 27 countries that contain factories with addresses, that are connected to the Mondial database by country. China contains the most factories with a number of 479, which is by far more than to the other countries.

1.2) Using Address_Extended: Count the Factories per Country. Note, it has an aggregate function.

The result shows 63 countries that contain factories with addresses, that are connected to the Mondial database by country, province and city. Here we can see Vietnam containing the most factories with a number of 497, closely followed up by China with 414.

1.3) Factory count per country for Address_Mondial and Address_Extended combined. Note, it has an aggregate function.

The result shows a variety of 65 countries for factories in our database. China with 893 and Vietnam with 497 factories have by far the most factories compared to the rest. From this we can conclude that a lot of sport product manufacturing is located in east asia.

Zimpfer

2) In this Section information about countries containing factories, focusing on EU and non-EU countries is explored.

2.1) All information for all EU member countries, with their country, organization and type.

$$\pi_{*}(\sigma_{organization="EU" \wedge type="member"}(ismember)$$

The result shows the 27 EU countries, which we will need for further queries.

2.2) The EU countries and their factory count per country. Note that Factory_Count_Country is a view.

$$\pi_{country\_code,factory\_Count\_All}(\sigma_{organization="EU" \wedge type="member"}(Factory\_Count\_Country \bowtie \text{ismember}))_{country\_code \rightarrow EU\_Countries}$$

The result shows 16 out of the 27 EU countries for factories producing sport goods.

2.3) The Non-EU countries and their factory count per country. Note that Factory_Count_Country is a view.

$$\pi_{country\_code,factory\_Count\_All}(Factory\_Count\_Country \triangleright \pi_{country\_code,factory\_Count\_All}$$
$$(\sigma_{organization="EU" \wedge type="member"}(Factory\_Count\_Country \bowtie \text{ismember}))_{country\_code \rightarrow EU\_Countries})_{country\_code \rightarrow Non\_EU\_Countries}$$

The result shows 49 out of all 63 countries producing sport goods in non-EU countries. By this we can conclude that sport product brands like to relocate their production outside of the EU.

2.4) The factory sum of all EU countries and their average GDP. Note, it has an aggregate function.

The results shows a total of 113 factories in the EU with an average GDP of 810636.

2.5) The factory sum of all non-EU countries. Note, it has an aggregate function.

The GDP was left out in this query, because it would give a distorted information compared to EU countries having similar GDPs and because most production takes place in China and Vietnam outside the EU. The result shows a total of 2704 factories in non-EU countries, which supports the argument of sport product brands liking to produce outside of the EU.

2.6) The brand's, China's and Vietnam's: brand or country, code, GDP, service, industry and factory count. Note that Factory_Count_All is a view.

$$\pi_{brand\_name,code,gdp,service,industry,factory\_count\_all}(Brand \bowtie \text{economy} \bowtie \text{Factory\_Count\_All})_{brand\_name \rightarrow brand\_or\_country} \cup \pi_{country,country,gdp,service,industry,factory\_count\_all}\left(\sigma_{country="CH"}(economy)\right) \cup \pi_{country,country,gdp,service,industry,factory\_count\_all}(\sigma_{country="VN"}(economy))$$

Based on the information, we can see that companies making sports products are mainly in Germany and the USA. The USA has the biggest GDP, does a lot of services, not as much manufacturing, and doesn't have many factories. Germany has less money than China, but it does a lot of services, even a bit more than the USA, and has the fewest factories.
China and Vietnam are similar because they do about the same amount of services and manufacturing, it's almost equal. China has more money than Germany but less than the USA, and it has a lot of factories. Vietnam has very little money compared to the other countries and almost half the factories of China.
Since China and Vietnam have lower GDP companies might pay less to workers there. This makes it cheaper for companies to move their production to these countries. Also, because these countries are good at manufacturing, it's more efficient for companies to make their products there.

Zimpfer

3) In this Section information about cities and the city's population containing factories is explored.

3.1) How many people of a population in a city might work in the factory with the same city? Worker_Amount_Range is used to compute the relation worker to city population. Return the factory name, city, the worker upper bound, city population, computation and brand name.

The aggregate function of ORDER BY is left out in the relational algebra. This will be only an indicator, because some workers might commute to work. Note that Worker_Amount_Range (in %) is limited to VF Corporation and Puma. Also note that we are optimistic in our calculations by using the upper bound instead of the lower bound.

$$\pi_{Factory\_Name,City-Name,Worker\_Amount\_Range.Worker\_Amount\_Upper\_Bound,City.Population,(Total\_Workers\ /\ City.Population)*100,Brand\_Name}$$
$$(\sigma_{Worker\_Amount\_Upper\_Bound \neq NULL}(Factory \bowtie Brand \bowtie \text{Address\_Mondial} \bowtie \text{City} \bowtie \text{Worker\_Amount\_Range})$$
$$City.Name \rightarrow City\_Name, Worker\_Amount\_Range.Worker\_Amount\_Upper\_Bound \rightarrow Total\_Workers, City.Population \rightarrow City\_Population,$$
$$(Total\_Workers\ /\ City.Population)*100 \rightarrow Worker\_to\_Population\_Relation\_in\_Percentage$$

The result is limited to VF Corporation, because Puma didn't provide data for province. Therefore no queries with Mondial's city table is possible due to the composite key of city, needing data for city, province and country. Note that only Addresses with a connection to Mondial appear in the result.

3.2) This is the same query as 3.1, except it uses Total_Worker_Int instead of Worker_Amount_Range to get results of Nike and Adidas. These brands are limited to Worker_Amount_Int.

Adidas is showing dominant results in worker to population relation. Note that only Addresses with a connection to Mondial appear in the result which could be a reason for more Adidas results than Nike.

4) In this Section information about factories having female and migrant workers is prepared.

4.1) All the factory's related information by name or number instead of the IDs.

$$\pi_{Factory\_ID,Factory\_Name,Brand\_Name,Product\_Type,Parent\_Organization,Supplier\_Group,Total\_Worker\_int,Female\_Worker\_int,Male\_Worker\_int,Migrant\_Worker\_int,Line\_Worker\_int,\\ Female\_Worker\_Percentage,Migrant\_Worker\_Percentage,Worker\_Amount\_Lower\_Bound,Worker\_Amount\_Upper\_Bound,am.Address,ae.Address}\\ (Factory \bowtie Brand \bowtie Product\_Type \bowtie Parent\_Organization \bowtie Supplier\_Group \bowtie Worker\_Amount\_Int \bowtie Worker\_Amount\_Percentage\\ \bowtie Worker\_Amount\_Range \bowtie Address\_Mondial \bowtie Address\_Extended)_{am.Address \rightarrow Address\_Mondial, ea.Address \rightarrow Address\_Extended}$$

The result shows the factory's information in a more comprehensive way.

4.2) The factory's ID, name, female and migrant percentage number, excluding rows of null. Worker_Amount_Int being computed in percentage, excluding rows of null. The ROUND function is left out in the Relational Algebra. Note that Factory_View is a view.

$$\pi_{Factory\_ID,Factory\_Name,Female\_Worker\_Percentage,Migrant\_Worker\_Percentage}(\sigma_{Female\_Worker\_Percentage \neq NULL \wedge Migrant\_Worker\_Percentage \neq NULL}\\ (Factory\_View)) \cup_{ALL} \pi_{Factory\_ID,Factory\_Name,(Female\_Worker\_Int * 100/Total\_Worker\_Int),(Migrant\_Worker\_Int * 100/Total\_Worker\_Int)}\\ (\sigma_{Female\_Worker\_Int \neq NULL \wedge Migrant\_Worker\_Int \neq NULL}(Factory\_View))$$

The result shows percentage values for the 2190 factories instead of total 2817 factories being provided in the database. The reason is not provided data by the brands (null values) for some factories.

Zimpfer

5) In this Section information about factories and their female worker amount is explored.

5.1) The factory's ID, name and where female worker percentage is 95 % or higher. Excluded is Factory with the ID 674, because the brand provided false data. Note that Factory_View is a view.

$$\pi_{Factory\_ID,Factory\_Name,Female\_Worker\_Percentage}(\sigma_{Female\_Worker\_Percentage \geq 95 \wedge Factory\_ID <> 674}(Factory\_Percentage \bowtie \text{Factory\_View})$$

The factory with the ID 674 would give an output of 393% female workers, which is impossible. Checking the data of Adidas, for the factory "Can Sports Shoes Co., Ltd." they gave a total worker count of 1652 and a female worker count of 6493 which result in 393%. Therefore I excluded this factory. This is where we can see the importance of working with correct data to avoid false interpretations. You can find this mistake for the factory in row 36 of the original Adidas table.

5.2) The Address_Extended and Address_Mondial for the factory with the ID of 674.

$$\pi_{address\_m\_id,address\_e\_id}(\sigma_{Factory\_ID=674}(Factory))$$

The result shows an ID for Address Extended, which will be excluded in the next query.

5.3) For each country the average of female factory worker in percentage, where it is 50% or higher. Excluded is Factory with the ID 674. Note, it has an aggregate function. Displayed in the following format:
"This country: <Country_Code> has this avg % of females: <ROUND(AVG(female_worker_percentage), 4)> in sport product factories."

The result shows 54 countries having a 50% or higher average amount of female workers.

5.4) For each country the average of female factory worker in percentage, where it is lower than 50%. Excluded is Factory with the ID 674. Note, it has an aggregate function. Displayed in the following format:
"This country: <Country_Code> has this avg % of females: <ROUND(AVG(female_worker_percentage), 4)> in sport product factories."

The result shows 10 countries having a lower than 50% average amount of female workers. Therefore we can conclude that most countries have female workers dominating in manufacturing factories for sport products.

References

[1] https://lms.hs-pforzheim.de/pluginfile.php/400406/mod_resource/content/0/3_1_relational_model.pdf

[2] https://lms.hs-pforzheim.de/pluginfile.php/400407/mod_resource/content/0/3_2_relational_model.pdf

[3] https://lms.hs-pforzheim.de/pluginfile.php/400408/mod_resource/content/0/3_3_relational_model.pdf

[4] https://lms.hs-pforzheim.de/pluginfile.php/400409/mod_resource/content/0/4_1_relational_databases-sql_introduction.pdf

[5] https://lms.hs-pforzheim.de/pluginfile.php/400410/mod_resource/content/0/4_2_relational_databases-extended_queries.pdf

[6] https://lms.hs-pforzheim.de/mod/glossary/showentry.php?eid=1804

[7] https://stackoverflow.com/questions/21538082/any-difference-between-union-all-corresponding-and-outer-union-corresponding