

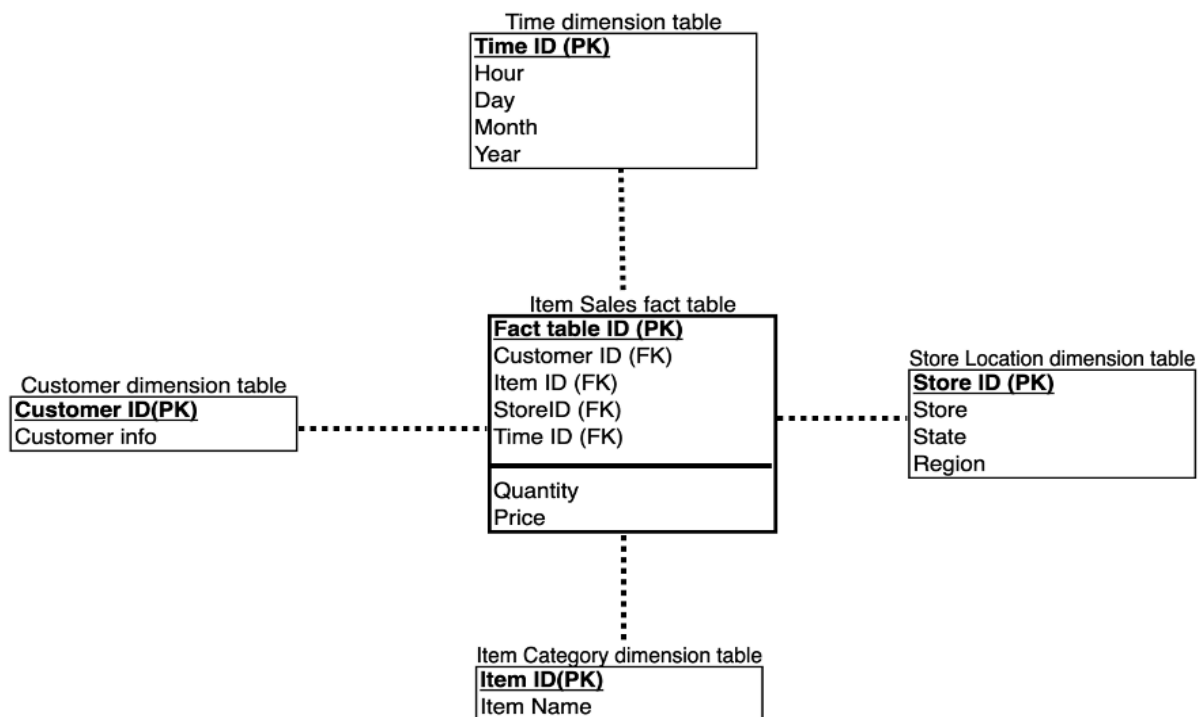
Amin Baabol
INFS 762
Assignment 3

Task 1

Data analysis sheet

Fact.....	Item Sales			
Measurements.....	Quantity of sold items, item price			
Dimensions	grain	hierarchies		
	Item Unit			
	Hour			
		Day		
			Month	
				Year
	Store			
		State		
			Region	
	Customer			

Star schema



Task 2:

OLAP with Cognos

1.A screenshot of a new tab you created that includes Figure 8 (on page 22 of the tutorial).

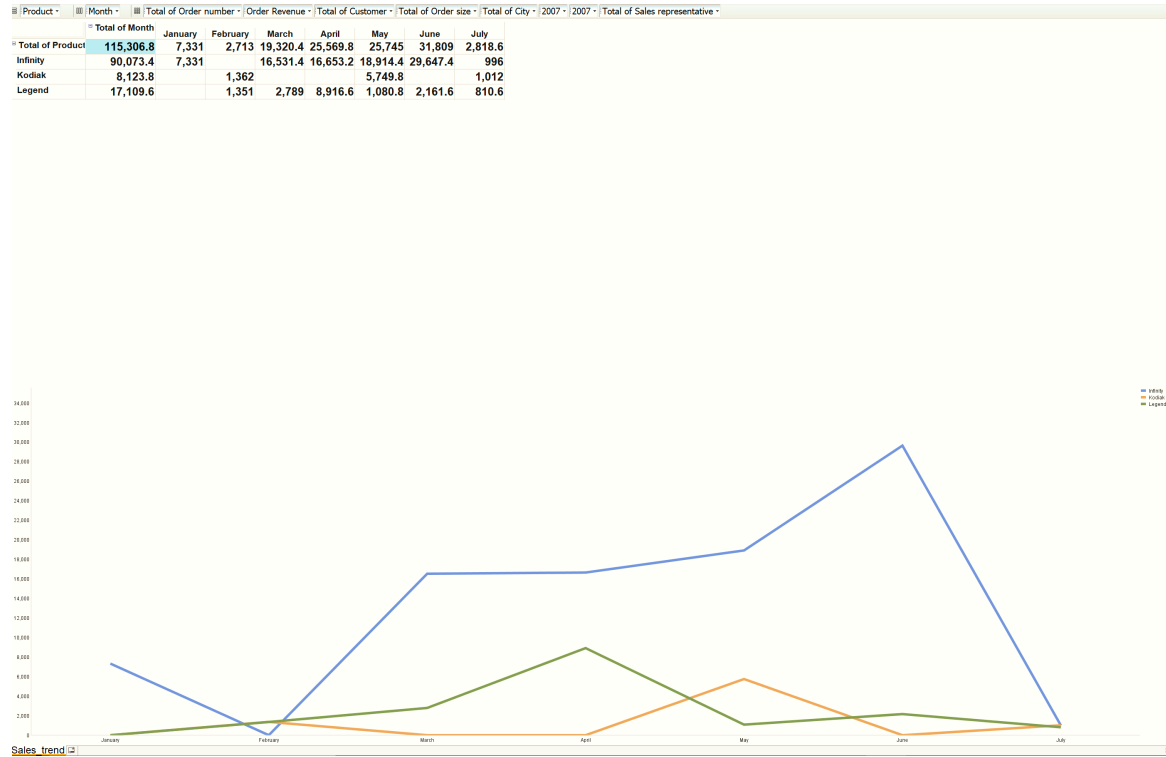
All Products ▾ All Customers ▾		Orders Measures ▾	All Order numbers ▾ All Months ▾ All Order sizes ▾
		Order Revenue	
All Products	All Customers	\$115,307	
	Golf's'us	\$7,518	
	Kanga Kampers	\$9,765	
	The Marketplace	\$1,016	
	Fachgeschäft Müller	\$32,139	
	Grand choix	\$5,371	
	Falcon Outfitters	\$52,379	
	Sportworld	\$7,118	
Infinity	All Customers	\$90,073	
	Golf's'us	\$6,506	
	Kanga Kampers	\$6,241	
	The Marketplace	\$1,016	
	Fachgeschäft Müller	\$24,768	
	Grand choix	\$3,750	

2.A screenshot of the new table that includes Figure 9 (on page 22).

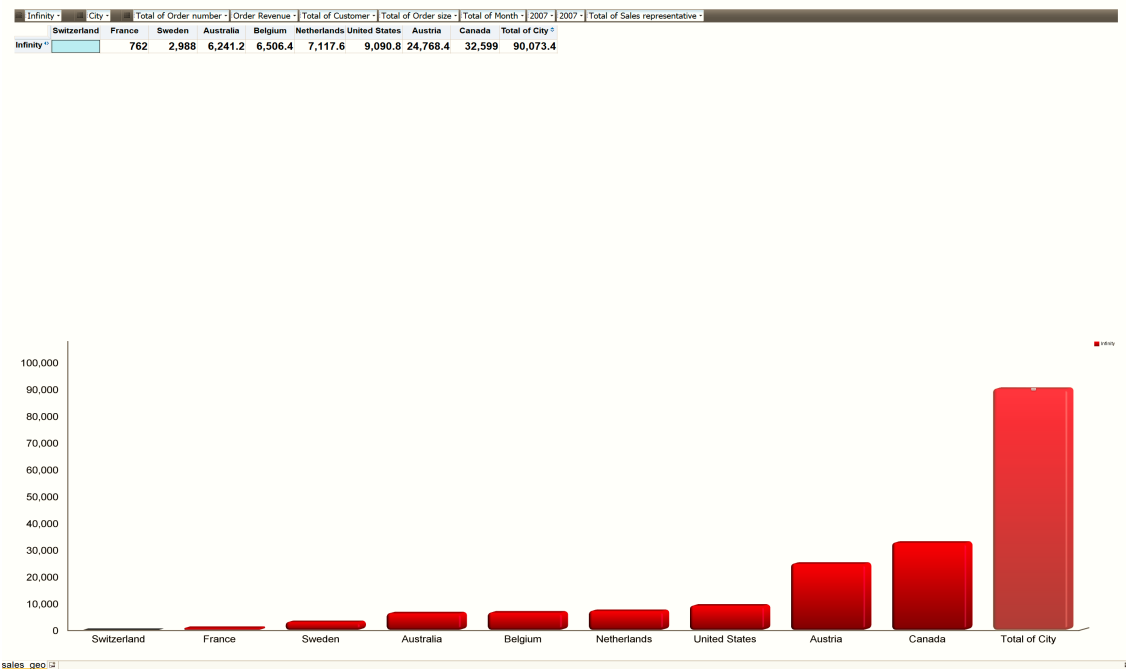
Products ▾		Orders Measures ▾		Month ▾	All Order numbers ▾		All Customers ▾	All Order sizes ▾	
		Order Revenue							
		▾ All Months	January	February	March	April	May	June	July
All Products		\$115,307	\$7,331	\$2,713	\$19,320	\$25,570	\$25,745	\$31,809	\$2,819
Infinity		\$90,073	\$7,331		\$16,531	\$16,653	\$18,914	\$29,647	\$996
Kodiak		\$8,124		\$1,362			\$5,750		\$1,012
Legend		\$17,110		\$1,351	\$2,789	\$8,917	\$1,081	\$2,162	\$811

3.OLAP analysis

You need to show a line graph like the following. As you can see the sales revenue of infinity is much higher than other products, but we see two significant drops of its sales in Feb and July. Please submit a screenshot of the tab.

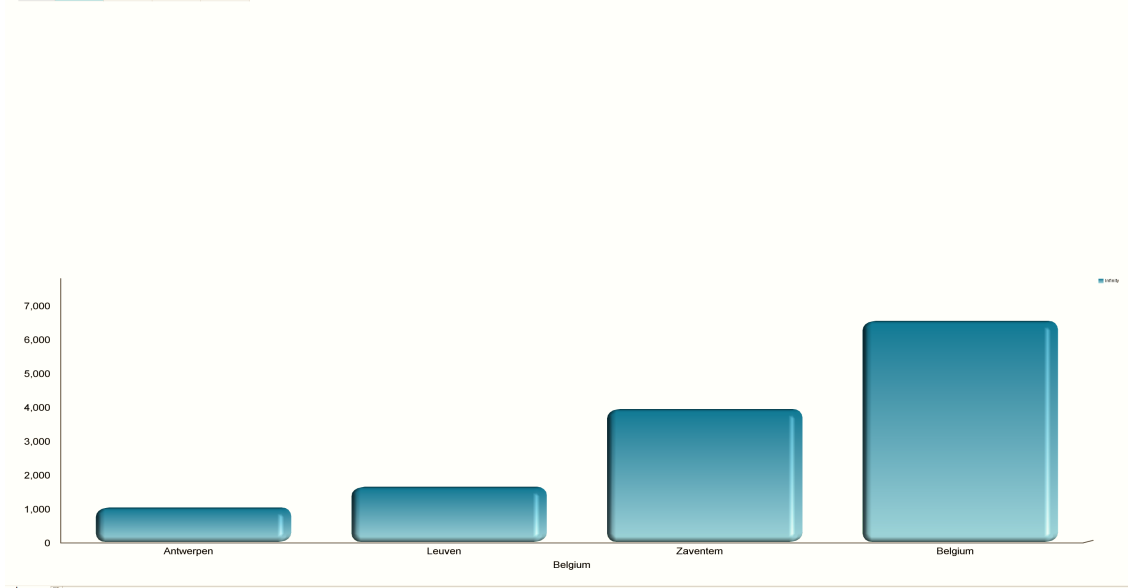


3. Please submit a screenshot of the tab. Please show me the sales revenue of infinity in different cities in Belgium and sort the cities according to the sales revenue and submit a screenshot of the tab (hint: you want to use the drill down function).



City	Infinity
Total of City	90,073.4
Canada	32,599
Austria	24,768.4
United States	9,090.8
Netherlands	7,117.6
Belgium	6,506.4
Australia	6,241.2
Sweden	2,988
France	762
Switzerland	

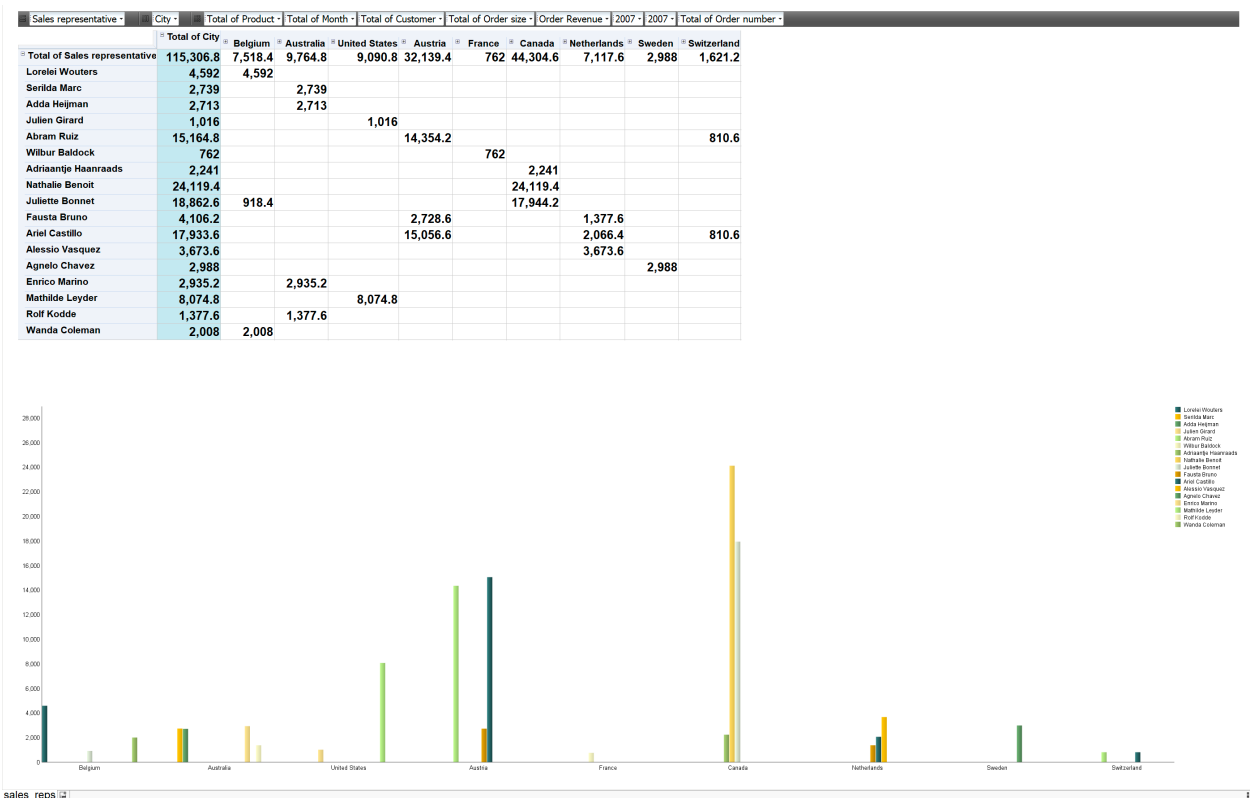
Infinity	Belgium	Total of Order number	Order Revenue	Total of Customer	Total of Order size	Total of Month	2007	2007	Total of Sales representative
Antwerpen	Leuven	Zaventem	Belgium						
Infinity	996	1,607.2	3,903.2	6,506.4					



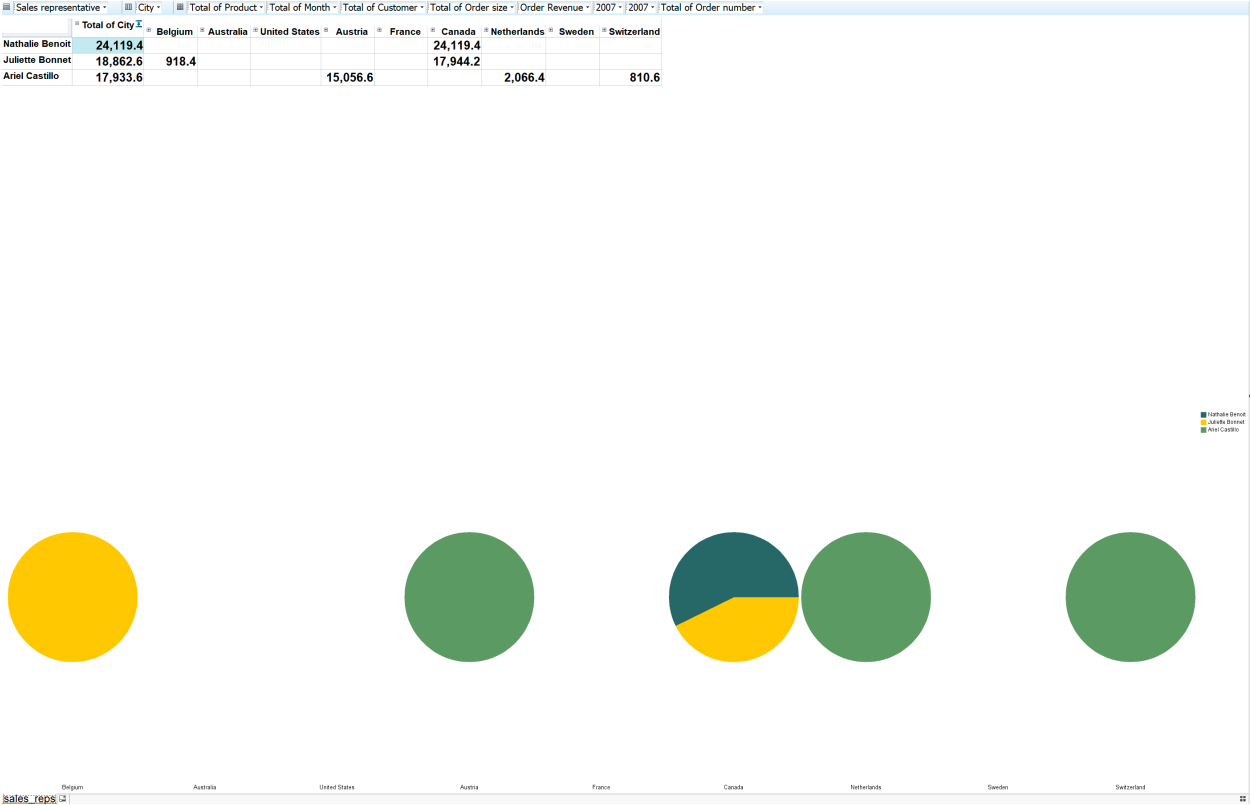
sales_geo [2]

	Infinity
Belgium	6,506.4
Belgium > Zaventem	3,903.2
Belgium > Leuven	1,607.2
Belgium > Antwerpen	996

4(1). Please create a new tab called “sales_reps”. Create a cross tabular table with the sale representatives on the x-axis and the countries on the y-axis and submit a screenshot. You don’t need to change the diagram below the table.



4(2). Please display the top 3 sales representatives in terms of total sales revenue of all products. You need to create pie graphs for these 3 representatives. Each pie graph shows the sales revenue of one sales representative in different countries. Please submit a screenshot of the tab. Please do not use the “sort” function. Only the top 3 representatives can be shown in the cross-tabular table.



Task 3. Short-answer questions:

- 1) What are the differences between data warehouses and OLTP systems (at least 3 differences)?
 - a. An OLTP system is designed to capture the day-to-day operational data for a short period of time. On the other hand, a data warehouse is designed to contain all of the historical operation data of an organization for a long time.
 - b. Records are frequently updated in an OLTP systems compared to a data warehouse where data is almost never over-written in order to maintain consistent future reporting.
 - c. Writing queries to read information from a data warehouse can get complicated and long as oppose to the OLTP's small transactions.
 - d. Data warehouses are mostly used by business decision makers, while OLTP is operated by clerical users.
- 2) How is a star schema different from a snowflake schema? What are the advantages and disadvantages of star schemas vs. snowflake schemas?
 - a. In star schema, the dimension tables are denormalized as oppose to snowflake schema where dimension tables are normalized.
 - b. Advantages of star schema include:
 - Simpler structure
 - Fewer tables
 - Faster query execution
 - c. Disadvantages of star schema include:
 - Higher potential data redundancy
 - It is not so flexible.
 - d. Advantages of snowflake schema include:
 - Efficient space storage
 - Low potential for data redundancy.
 - e. Disadvantage of a snowflake schema include:
 - It requires complex
 - Slower query processing time to due to the dimension tables normalization.
- 3) We can implement a data warehouse using either a relational database or multidimensional database. What the advantages and disadvantages of each of the methods?
 - a. Advantages: Multidimensional
 - Efficient storage and retrieval of large volume data
 - Enhanced convenience in viewing and analyzing data
 - Optimized for DW and OLAP applications
 - b. Disadvantages: Multidimensional
 - Not ideally suited for unrelated data types
 - Relies on relational database
 - c. Advantages: Relational database
 - Easy to use since it's stored in tables with rows and columns
 - Security feature can be implemented
 - High degree of flexibility
 - Optimized for DW and OLAP applications
 - d. Disadvantages: Relational database
 - Slower performance with increasing dimensions
 - Takes more storage space especially with normalization

- 4) In the LinkedIn Learning course “Implementing a Data Warehouse SQL Server 2019” by Adam Wilbert – you are required to watch the course in week 12, the author talked about how to develop *columnstore indexes*. Please use your own language (please do not directly copy/paste answers you find online) to briefly describe what columnstore indexes are and why they are useful for data warehouses.

Columnstore index is a method where the content of each column in a table is stored in its own individual page rather than grouping column data by row. This reduces the load on the disk by taking up less disk storage. One of the benefits of columnstore index is that it significantly improves query execution time which is suitable for processing large data tables in data warehouses faster.