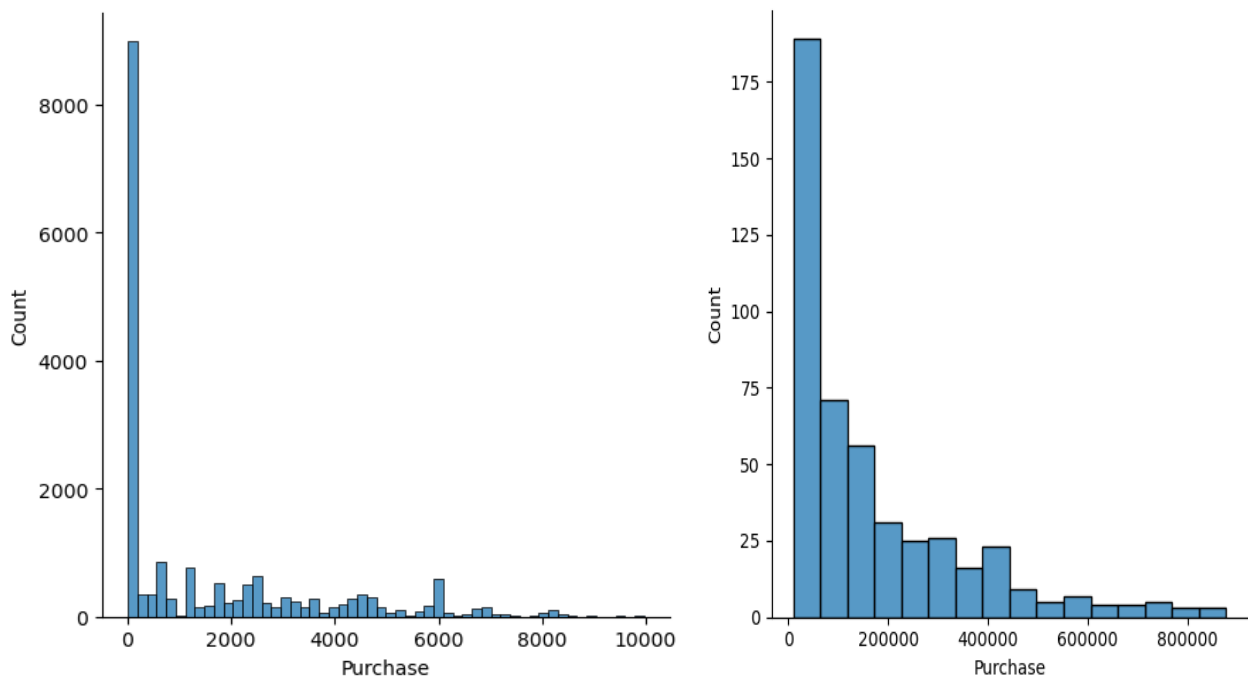


## AdventureWorks2022: an analysis.

AdventureWorks2022 is a database created to collect data about the activity of a shop selling bicycle and bikes equipmentmen. It is a big database composed by numerous schemas , tables and sales columns. There are 18 schemas including system schemas; the data stored regards important aspects of the shop's business starting with details about the products sold, specifications about the customers and of course details about the sales. The amount of data is broad and not always easy to manage. The data spans over a period of about four years and provide pieces of information regarding sales in 3 different continents and 10 different macroregions. Moreover one can father info about the state and the city where a customer lives. Though not complete it is possible to retrieve data not only customers, but also about sellers.

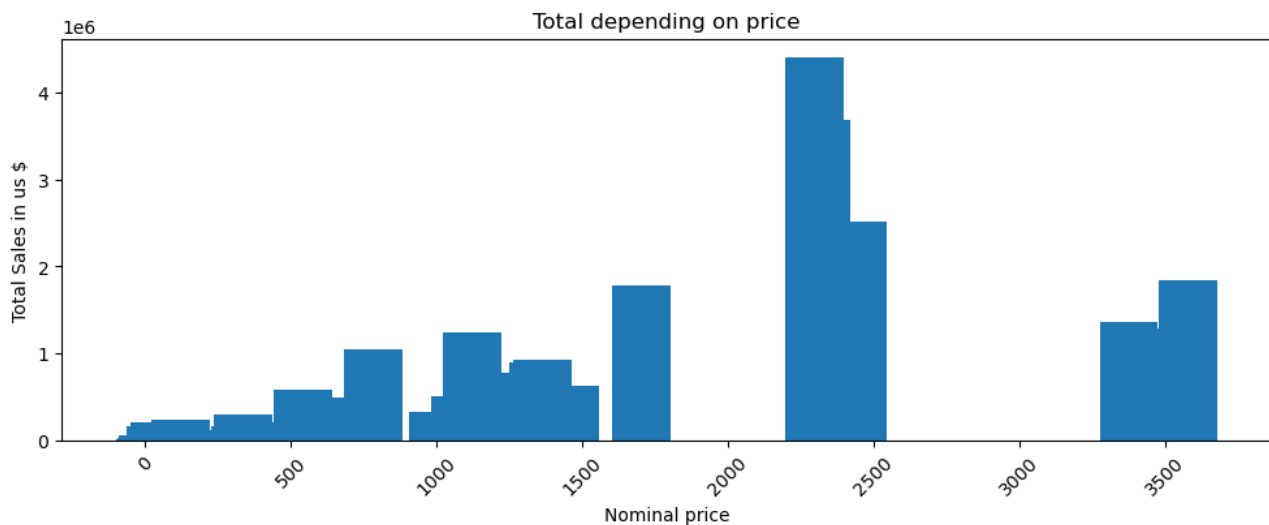
My analysis will focus on the main aspect of a shop: sales.

The database collects records about 19119 who have purchased something from the shop. On average a customer spent 5750 \$. The standard deviation is about 7 times bigger than the average; this fact tell us that the collected data are extremely spread and don't follow a normal distribution.



The images clearly show that big purchases are rare and the vast majority of operations regards small amounts! Purchases over 80000 US\$ are quite rare while on the other hand there are tens of thousands operations that sum up to less than 100 \$.

Another interesting data to look over is the distribution of sales depending on the unit price of a product.

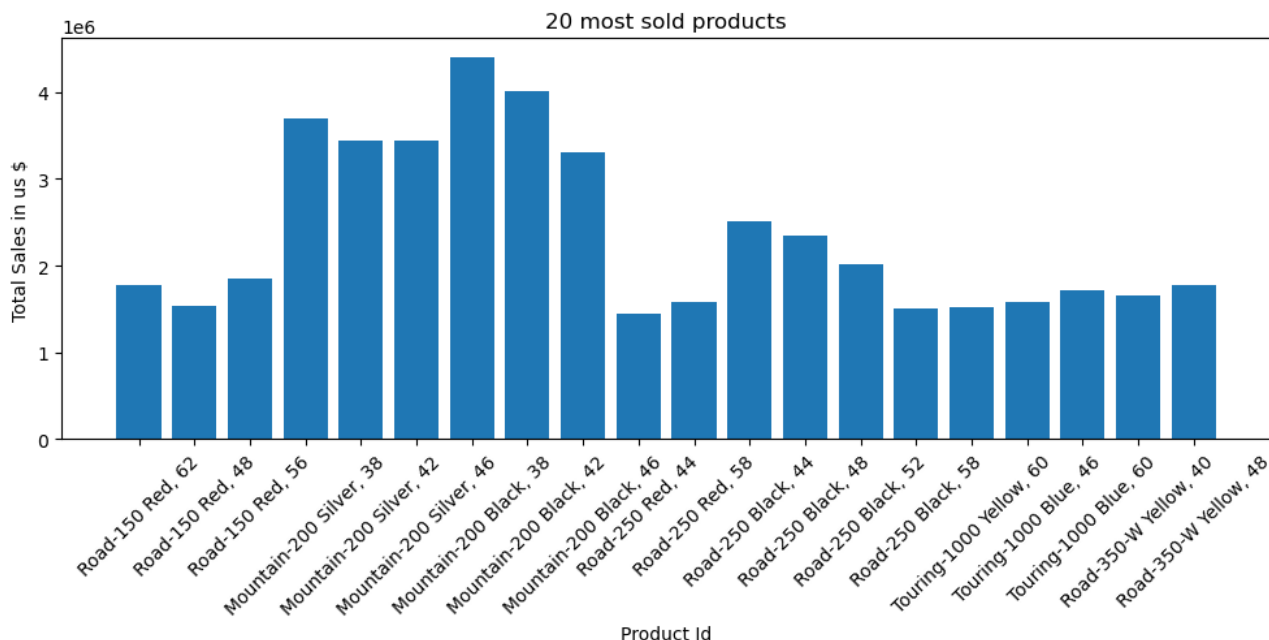


This distributions resembles a bit more a normal one: the more a product cost the more it generates high revenues. When the price Unit is bigger than 2500 then the revenues start to fall.

Among the data I collected this is the one that I chose to find an interval that contain the mean with 95% possibilities.

This is the result of the operation: (array([ 847.39438782, 653.60661117, 328065.46981023]), array([ 867.88380767, 875.69989259, 497847.17229419]))

We move on to other interesting aspects: which are the 20 most sold items?

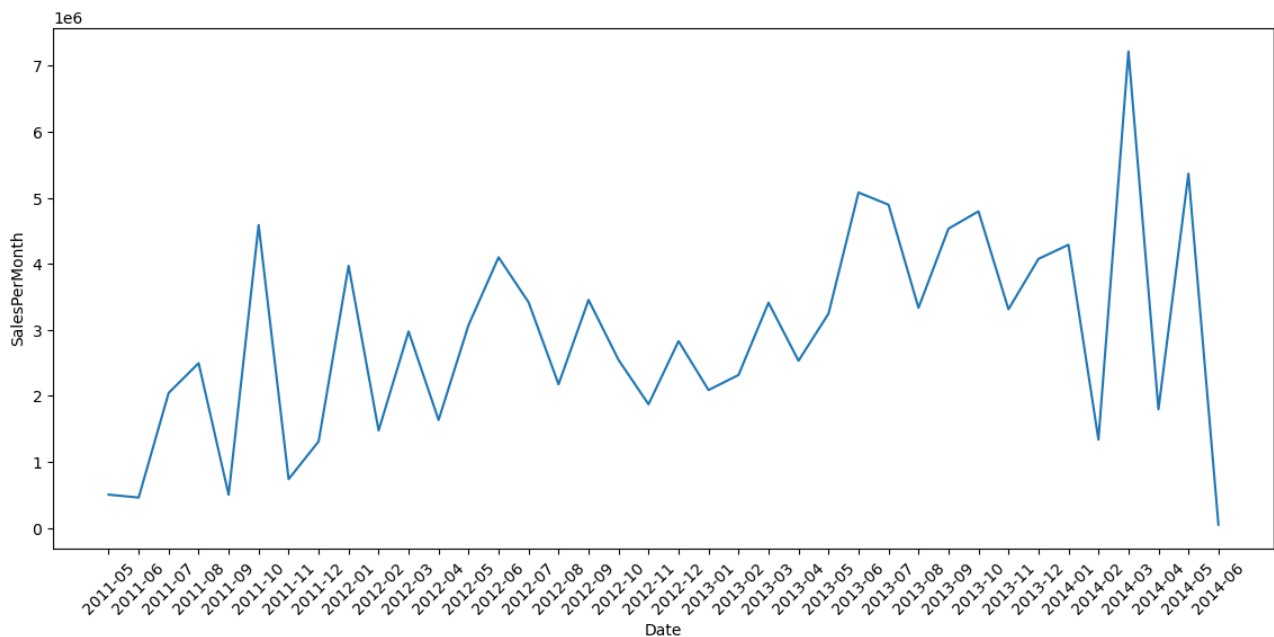


The image above shows the names of the 20 most sold products and the amount of the revenue they generate. The top selling bike generated sales worth than 4 millions US dollars.

Those 20 products alone sale for more then 47000000 \$ (47120773.250288)

## Monthly sales:

Talking about the monthly sales it is quite evident that the sales are not constant during the year

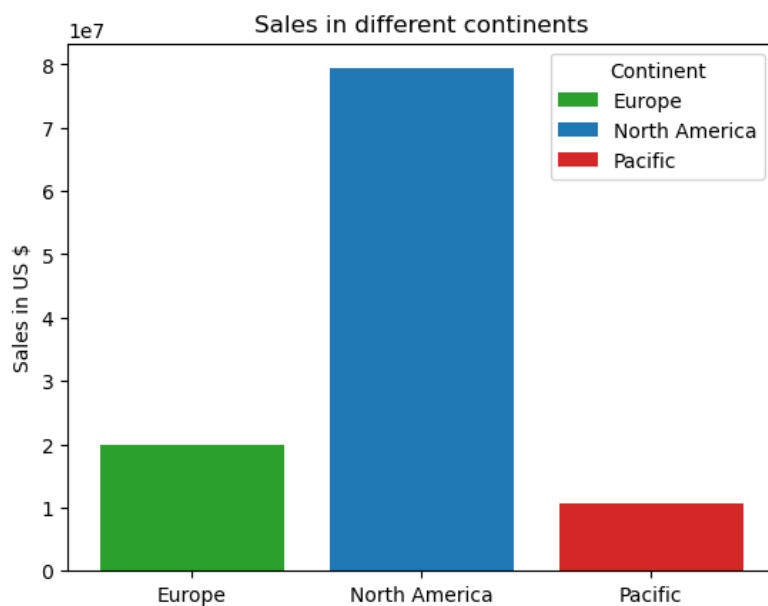


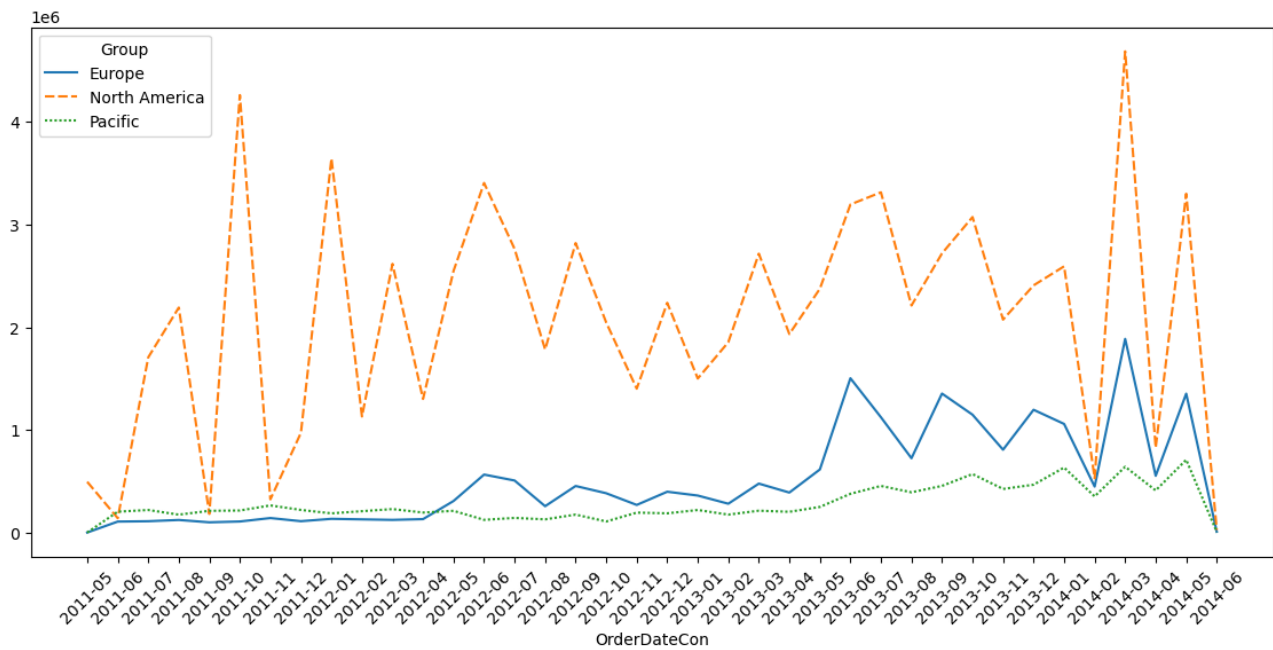
Nevertheless it is evident that there is a positive trend and the sales are growing. At the beginning of 2014 the monthly sales reached their peak.

## Geographical areas:

The analysis approaches its end with some interesting stats regarding different markets.

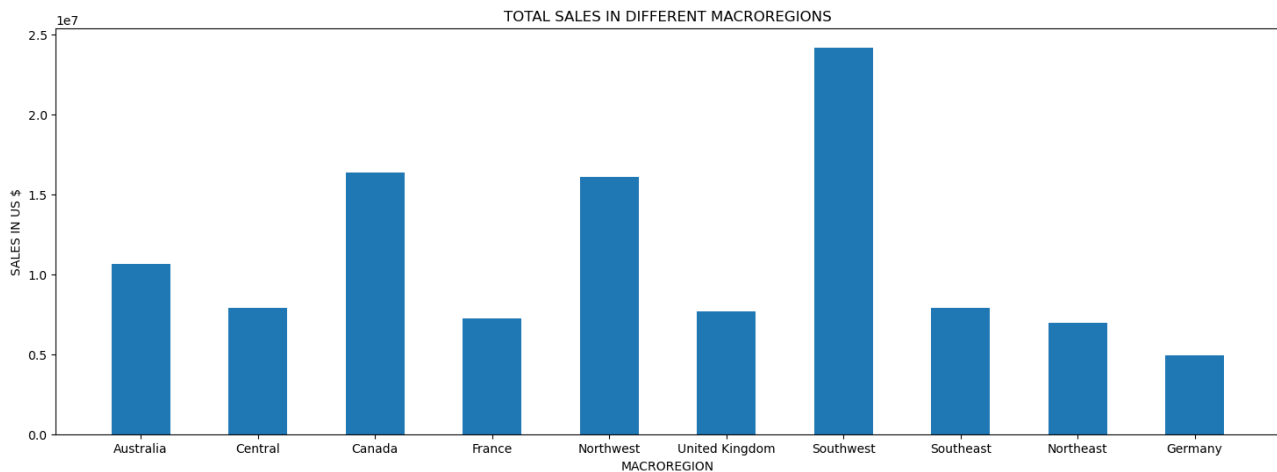
North America is the biggest market and its sales alone are more than double the sales of Europe and Oceania combined. Talking about the monthly sales it is interesting to see that while in Oceania and Europe the sales are more stable, in North America they alternate high levels to sudden falls. Only by the end of the considered period Europe and Oceania have a similar trend; this might be due to bad global markets conditions



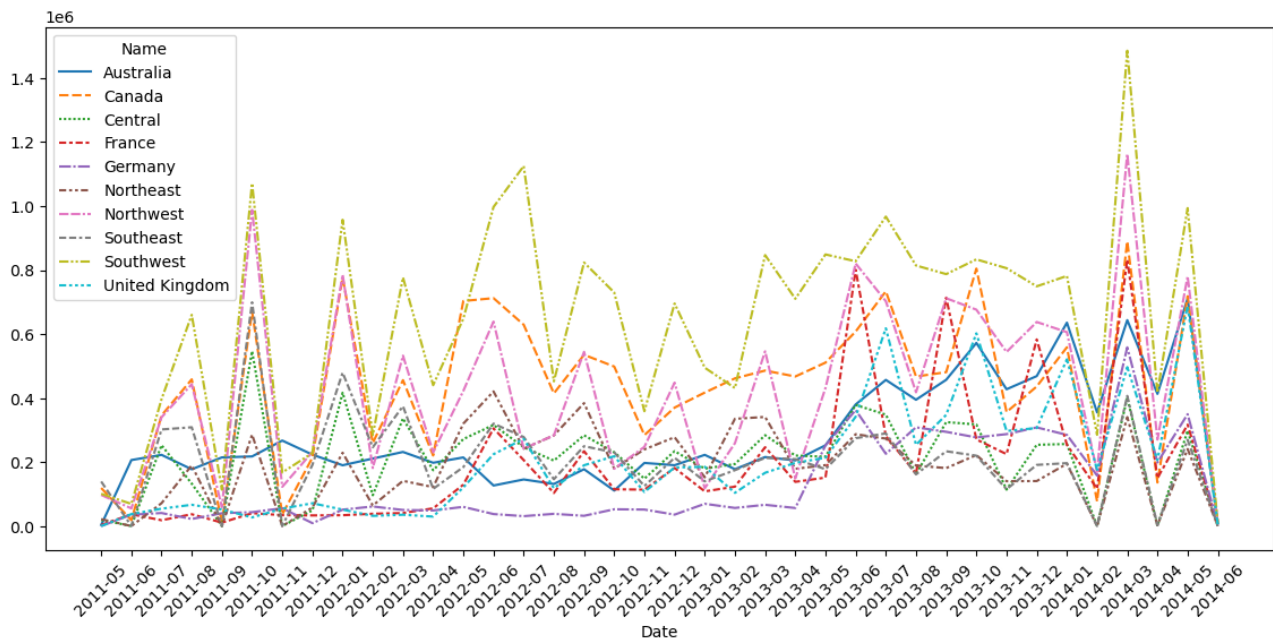


## Macroregions:

Lastly, let's have a look at the sales in ten different Macroregions.



United States and Canada confirm to be the most important market for these products. It is interesting though to notice that the north and the south have major differences in terms of sales and also the east and the west differ considerably. In Europe the biggest market is the UK where sales amount to more around 8 millions US \$; France is quite close.



This last picture just confirms what we saw monthly sales by continent.

## Considerations:

It is quite clear that the main market is North America and although comparable inhabitants and a and GDP per capita Europe is still considerably behind: Germany in particular has not reached its full potential. Talking about the variations occurring month after month it is unfortunately hard to understand what causes these fluctuations. A deeper understanding of the society could help our research. Lastly, though the range of products is quite wide it is undeniable that the main source of income are small product with a very low cost

## Executive Summary:

- 1 Different areas within the same continent generate significantly different revenues
- 2 Different continents generate different revenues
- 3 Germany and the NorthEast are two area where improvement are needed
- 4 An investigation is needed to understand why these rich areas are underperforming
- 5 Asia is completely absent; Japan and China are two of the most important markets in the world
- 6 An investigation is needed to understand why monthly sales differ so much
- 7 The sale trend is positive and shows an increment in sales
- 8 The price seems to have a huge impact on a customer's decision process

## About this work:

The main challenge was to understand the database actually. Many times similar column names generated confusions; it took a while to understand where to find the actual sales. Moreover different tables gave different results; this is for instance the case of the sales in the current year.

On a personal level the main challenge was to come back to this work after many months. Tips to myself? Don't waste time!