

DA3 – 2nd Group Assignment

Yetkin, Cagdas

Gomes, Nuno

The aim of this assignment is to assess a possible causal relationship between CEO's age and company's performance.

We've started our analysis with a csv. file called "bisbode_all.csv".

Our Database

Initially some of the variables seemed quite useful in a context like this, in which we wanted to understand if young CEOs perform better than old ones, but as our analysis deepened we started facing some difficulties and limitations regarding their dependencies.

A few observations were missing on key variables for our analysis. We've dropped and kept some of the observations in which this was happening according to each case and our evaluation. As the report develops we will explain what was done to clean the data in order to come up with a workable dataset, and we will explain what was the rationale supporting our decisions.

When checking the data, aside from the missing values, we've noticed that some of the observations made no sense in this context – e.g. negative age or children as CEO or absolute negative sales.

The data comprehended several years and distinct periods. We've decided to develop our analysis considering 2015 as the reference year, and using 2014 for comparison reasons in order to understand potential evolution/changes. 2015 was the year for which there was a better ratio in terms of missing observations on the crucial variables.

The dataset provides us with observations related to two main industries, being them Manufacturing (we've considered "C Manuf_auto" and "C Manuf_equipment" to be under the same industry classification) and Hotels & Restaurants ("I hotel and restaurant").

We've started by dropping all the observations that were not related to 2014 and 2015. Furthermore, we've noticed that the data was covering different periods, of different extents, therefore this heterogeneity could bring us some challenges when trying to directly compare both years and variations. That said, it was our decision to only keep the observations that were covering the period between 01-01-2014 and 31-12-2014 for 2014 and 01-01-2015 and 31-12-2015 for 2015. By doing this it was our understanding that we could directly compare the values

assuming a regular accounting/fiscal year knowing that this period refers to the whole year. In this case the fiscal year is matching the civil year.

Already with the time frame established, we've narrowed our criteria by deciding that only companies for which there were financial figures for a period equal or higher than 270 days should be considered ("balsheet_length". This way we've excluded the companies for which we only had a few days or months of data, making it difficult and not representative of year's performance.

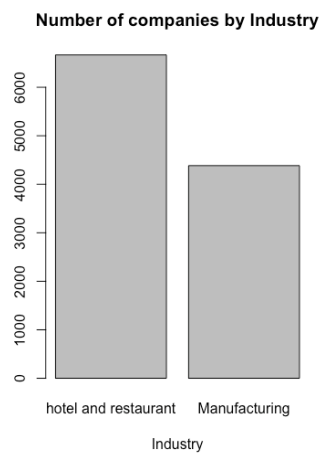
Considering that CEO's Age is crucial to our analysis we've excluded the observations for which there were no indications of birth year (birth_year). Without this input we wouldn't be able to calculate the Age. Still working on the age variable we've decided to only consider CEOs older than 20 years. To us a young CEO is between 20 and 40 years old, being that an old CEO is more than 40 years old – a binary variable was created "young".

Given that our aim is to understand a possible relationship between CEO's age and company's performance, we've decided to only consider the observations in which CEO was at that position for more than, or at least, 172 days (inoffice_days - the data considers that 1 year is equal to 364 days). It is our opinion that in a period of less than 6 months the impact of CEO's actions/implementations/cultural changes are still very restrained and hardly reflected on company's performance.

Keeping in mind its representativeness and data accuracy, we've decided to proceed only considering companies whose sales figures were comprehended between 100k EUR and 100M EUR. By doing this we were considering SME (Small and Medium Enterprises), therefore excluding micro and big companies. When we excluded the microenterprises we also took into account that usually, for these kind of companies, financial statements are not easy to get and they are usually less accurate.

After the above mentioned adjustments we ended up with a database that covers 2014 and 2015 from January to December. The companies that are portrayed include cases in which there has been only one CEO as well as cases in which the company had had several CEOs (e.g. 7).

On table 1 we can see that we have more observations from Hotels and Restaurants (6670) than from Manufacturing companies (4390).

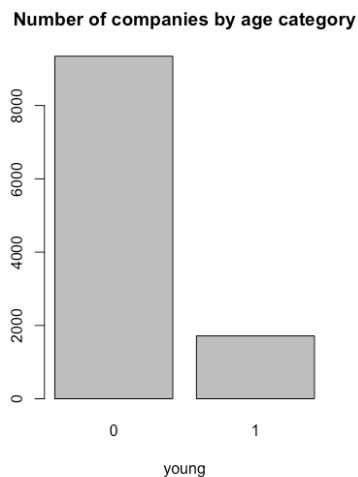


Industry	#
Manufacturing	4390
Hotel & Restaurant	6670

Table 1. Number of companies by industry

Plot 1. Number of companies by Industry

We can notice that the majority of the companies in our dataset are managed by old CEOs (**84%**).

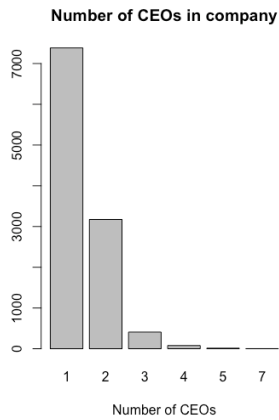


Age Category	# companies	%
Old	9335	84.48%
Young	1715	15.52%
total	11050	100.00%

Table 2. Number of companies by age category

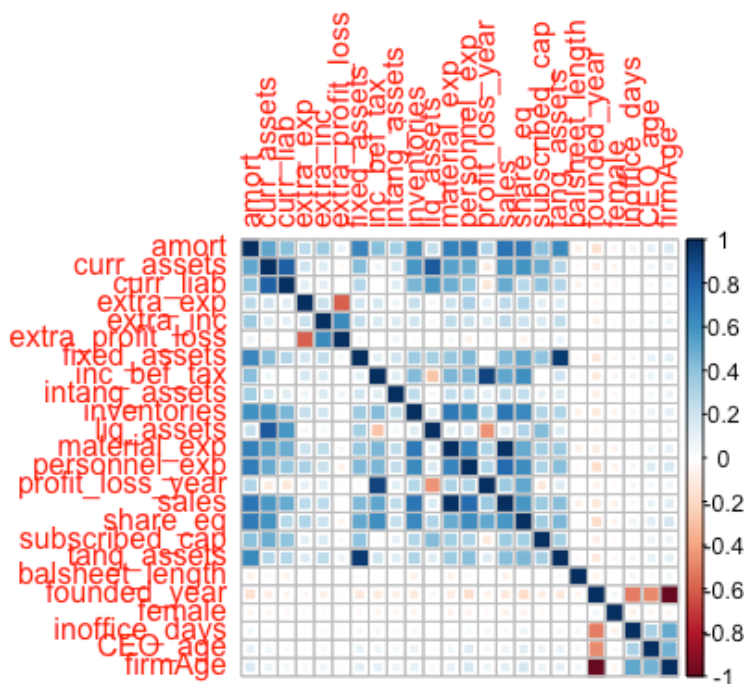
Plot2 – Number of companies by age category

At this stage we may state that we might face some issues regarding representativeness given that the number of companies managed by young CEOs are far less represented than companies led by old CEOs. We might argue that this database may not represent the population.



Plot3 – Number of CEOs per company

Furtherly we can state that most of the companies only met one CEO during this time period. After our initial adjustments the max number of CEOs drop from 15 to 7. In our analysis we are going to consider two scenarios: one in which the company only met a single CEO and another one in which the company was led by one or more CEOs.



Plot4 – Correlation matrix of numeric variables

1st Analysis – Several CEOs

Aside from all the adjustments previously described, initially we've decided to keep all the companies in which there were/had been several CEOs (at this point minimum was 1 and maximum were 7).

2nd Analysis – Only one CEO

For all of the companies that we were considering, there were available CEO's date of birth and in office days, but even so, limited to the database at our disposal, sometimes was not clear when there was the change in the position and if that change was from a young to an old or from an old to a young. That said, in order to cover both scenarios, we've decided to perform the same analysis this time for a database in which there were companies with only one 1 CEO.

Common ground

We were not sure what to expect from this analysis (segregating by number of CEOs) and if the data available would give us any special insight.

Our initial perception, not necessarily sustained by statistical analysis but more on empirical raw knowledge, was that usually when a company has several CEOs under a limited period of time that usually is due to some instability that usually reflects itself in less promissory results. We were also aware that are cases in which there is more than CEO in exercise (co-CEO scenario) or top management considered as CEO material.

After running some over these two subsets we've noticed that we were able to retrieve similar results with no noticeable differences.

Score

In order to evaluate CEO's performance we've tried to base our analysis in a scoring system, and for that we've decided to go a step further and analyze how CEO's age could have some sort of impact on company's level of expenses, profit/loss, liabilities and liquid assets levels. Basically, before considering these variables we wanted to understand if individually they would show any sort of correlation with age.

- Regression Expenses on age

To calculate the total of expenses we've considered extra expenses ("ext_exp"), expenses with materials ("material_exp") and expenses with personnel ("personnel_exp").

From this point on we've calculated how expenses in 2015 compare with expenses in 2014.

When we ran a regression of expenses on age, we found that it is not significant.

- Regression Profit/Loss on age

Similarly to what was done with expenses, we've calculated how Profit/Loss figures in 2015 compare with the ones in 2014 and the finding is that once again it is not significant.

- Regression Liabilities on Age

We've calculated how figures from 2015 compared with the ones from 2014 and then we ran a regression of Liabilities on Age. Once again we got similar results, it is not significant.

- Regression Liquid Assets on Age

The same was followed here, figures from 2015 were compared to the ones in 2014 and we ran a regression of Liquid Assets on Age. Not significant as well.

With this exercise we were trying to evaluate which variables to include and consider in a scoring system in order to evaluate company's performance. Previously we've seen that Sales and Age might have a relationship, but we wanted to include several other variables that, according to our understanding, would be useful when evaluating if a company is performing better or not. Although these variables might be useful when evaluating a company, without a proper context they might lead to wrong or incomplete conclusions.

An increase in expenses does not necessarily mean that the company is less efficient and it's management is not able to control the costs. Given the short scope period of this analysis it may fail to indicate that a company may be recently investing (extra expenses), producing more (expenses with materials) or hiring more than on the previous years (expenses with personnel).

Similar rationale applies to the Profit/Loss. Considering only a period of 2 years (2014 and 2015) and the information at our disposal, it is not easy and correct to draw a conclusion only based on Profit/Loss figures. As previously stated, a company may be aggressively investing and therefore the short-term results may be impacted and expected to be compensated on the long run (long run decisions are "ignored" here given the scope period). Other things might justify abrupt drops in profit, especially when there is a change on the CEO chair, as is the case of a "cold bath" – the new CEO might be "cleaning the house" and recognizing all the losses, devaluations, and

provisions at once, therefore attributing the results to previous management, this way starting his/her endeavor with “clean record”. Nonetheless, and as we were trying to assess, it could also be related to CEO’s age and her/his ability to manage, skills, inexperience, maturity.

An increase in liabilities is usually seen as something less positive, given that it directly impacts company’s solvability, investors will demand a higher premium and banks will demand more guarantees and access to credit will be constrained. Notwithstanding it may also be related to a period in which the company is investing, for instance, in a new facility in order to increase the production or it might have renewed the existing one in order to cope with new regulations or customer’s requirements, or might have increased the number of rooms or tables. If there is an increasing trend in liabilities for a long period that is usually a bad indication, but here we only have a static indicator on two years.

Usually an increase on liquid assets may be seen as something positive, given that we are able to cover the responsibilities related to them with their underlying value, with a remaining surplus. But even here we might want to consider a few scenarios that can make our analysis a bit more difficult.

Selling assets (sale and lease back or even a plain sale) directly impact this ratio and others, generating a decrease on the assets side and an increase on the revenue side (extra), that is reflected on period’s profit. Some analysts may see this as good given that there was an extra inflow, while others might argue that on the long run it means that a source of income and collateral eroded from company’s balance sheet. Assets that are registered under this rubric may hide considerable expenses in terms of maintenance, security, rotation and so on, that have direct impact on period’s results – P&L results.

When directing the discussion to how age affects a company’s performance, usually people refer some of the characteristics acquired/developed with as they get old and establish a relationship with a management style.

Young CEOs are said to be more knowledgeable regarding technology and the evolution that surrounds them, prone to take risks, disruptive, as well as more impatient and ambitious. On the other hand old CEOs are said to be more calm, more experienced, averse to risk (risk awareness) and more focused on stability.

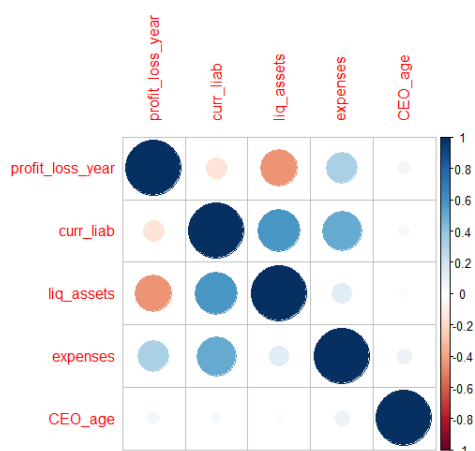
Both management styles are valid and each one of them has it’s pros and cons, depending on the context, industry, competition, and the moment in which the company is (if it is an underdog, star, cash-cow or even question mark on it’s industry, or if on the cycle it is expanding, consolidating, declining or mature). A complete and thorough analysis should consider all of this.

We've decided not to include on our analysis the region and the gender. That does not mean that we do not consider them theoretically important when justifying enterprise's performance.

We are aware, and open to consider the possibility, that usually there are differences that come with gender when referring to management style - women are said to be more accessible and personally attached than men, which could impact tougher decisions as downsizings and lay-offs. On the other hand women are said to be more organized than men, which might play an important role when dealing with the chaos of taking over CEO's position and dealing with a lot of decisions at the same time. Men are said to be more pragmatic than women, which might help the company to take the necessary actions on an increasingly competitive market in order to stay afloat.

The region may also have an impact on company's performance. Depending on the country, it's size, culture, natural resources, trade agreements, demography and so on, the regions may be organized/exposed in different manners and therefore some regions may be more prolific than others. One region might have a cluster in terms of manufacturing (auto manufacturing and components), therefore benefiting from scale costs and proximity to it's suppliers, while other region might be more exposed to Tourism based on natural resources, culture, architecture and so on. Other regions may have a limited exposure to any of those. To go further with our analysis regarding region we would need more information than the one available at this point.

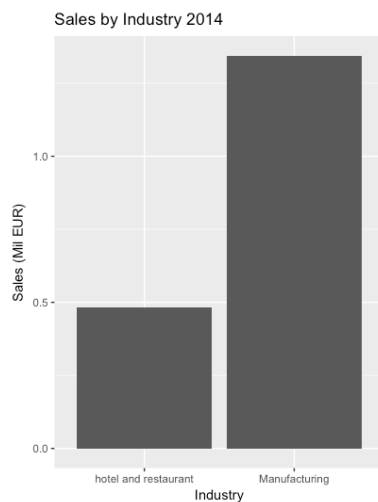
The data available was not enough to sustain any conclusion or corroborate any of this.



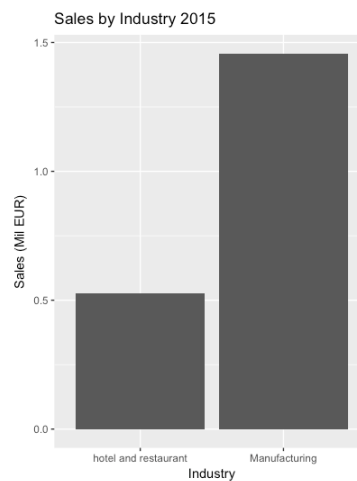
Plot5 – Correlation Matrix

Analyzing company's performance using sales

We've started by analyzing the sales volume in 2014 and 2015, calculating the change from one year to the other.



Plot6 – Sales (Mil EUR) by Industry – 2014



Plot7 – Sales (Mil EUR) by Industry - 2015

On Plots 4 and 5 we can see the average sales per Industry. We can easily notice that although on our database we have more observations from Hotels and Restaurants, on average the sales are higher on Manufacturing companies. Now it comes to reason that in Hungary there several automakers (e.g. Mercedes, Audi, Suzuki) with local production that they usually export (production lines) which positively drives the results of the manufacturing industry on this country. We could also mention other manufacturing facilities (lightning, electronic components to several industries) that are also represented in Hungary.

Regarding Hotels and Restaurants we should consider the fact that for the past few years there has been an increase on the number of Tourists coming to Hungary, specially to Budapest given that it is getting trendy, but also people that is coming to do business (given the number of international companies with office here), which may partially justify the increase in sales year after year.

There were cases with abnormal increases in sales that biased the results. It is not normal to have such an increase in such a limited period of time. In a different context we would revisit the initial databases from which the information was initially retrieve in order to assess if there was

any additional information that might justify such a spike, and additionally we would search for information related to these entities (for instance online or business directories), but for the sake of the current analysis and considering that these outliers are biasing the final results we've pondered and decided to exclude them.

We've decided to create a new variable in order to capture this difference in sales volume (2015-2014) and we ran a regression for this variable on age.

Statistic	N	Mean	St. Dev.	Min	Max
comp_id	5,469	152,999,117,280.000	137,514,190,590.000	1,002,029.000	464,105,013,248.000
sales.x	5,469	822,300.000	2,569,704.000	100,025.900	74,110,856.000
score.x	5,469	1.300	77.652	-457.254	5,711.265
CEO_age.x	5,469	50.541	10.514	21	89
firmAge.x	5,468	11.465	7.391	0	30
sales.y	5,469	894,677.200	3,029,668.000	100,040.700	89,543,536.000
score.y	5,469	1.530	77.686	-203.256	5,715.206
CEO_age.y	5,469	50.358	10.536	21	89
firmAge.y	5,468	12.465	7.391	1	31
salesdiff	5,469	0.114	0.330	-0.916	1.948
success	5,469	0.500	0.500	0	1

Table3– Summary

Initially a regression of “Sales Difference” on Age was ran and we were able to identify a possible relationship between them.

Log sales diff: $\text{Logsales}(2015) - \text{Logsales}(2014)$

Sales Difference: $(\text{Sales}2015 - \text{Sales}2014) / \text{Sales}2014$

Complementarily other regressions were ran in order to understand if other variables would proportionate a model that would better fit our aspirations.

	<i>Dependent variable:</i>			
		salesdiff	(log(sales.y) - log(sales.x))	
	(1)	(2)	(3)	(4)
young.y1	0.057*** (0.012)	0.031** (0.013)	0.031** (0.013)	0.035*** (0.012)
firmAge.y		-0.004*** (0.001)	-0.004*** (0.001)	
ind.yManufacturing			-0.005 (0.009)	
Constant	0.103*** (0.005)	0.161*** (0.010)	0.162*** (0.010)	0.058*** (0.005)
Observations	5,234	5,233	5,233	5,234
R ²	0.004	0.013	0.013	0.002
Adjusted R ²	0.004	0.012	0.012	0.002
Residual Std. Error	0.326 (df = 5232)	0.325 (df = 5230)	0.325 (df = 5229)	0.305 (df = 5232)
<i>Note:</i>				*p **p ***p<0.01

Table4— Regression Sales on Age and other variables

In all the regressions Age comes up as significant, being that it's coefficient is not considerably high. Additionally we can also mention the low R² coefficients, which at first glance indicate that our model is only able to explain a small fraction of the variation in sales between 2014 and 2015. Even when we've included some of the other variables available in the dataset, there was no considerable change on the explanatory power of our model.

A majority of the variation is left with no explanation. In a situation like this the analyst should consider the inclusion of other variables that could help to better understand the variance in sales. These variables could come from a micro or macro level. On a micro level we could consider the level of competition, local regulations (restrictions or incentives to open restaurants, hotels, factories), minimum wage law and so on. On a Macro level the analyst would have to consider the exchange rates (both for comparability reasons as well as for transfer pricing given that we may be dealing with international firms), price of oil (if manufacturing heavily depends on electricity), Regional crisis and so on.

<i>Dependent variable:</i>			
	success		
	(1)	(2)	(3)
young.y1	0.073*** (0.019)	0.036* (0.020)	0.032 (0.020)
firmAge.y		-0.006*** (0.001)	-0.005*** (0.001)
ind.yManufacturing			-0.053*** (0.015)
Constant	0.488*** (0.008)	0.572*** (0.015)	0.583*** (0.015)
Observations	5,234	5,233	5,233
R ²	0.003	0.011	0.013
Adjusted R ²	0.003	0.010	0.013
Residual Std. Error	0.499 (df = 5232)	0.497 (df = 5230)	0.497 (df = 5229)
F Statistic	14.615*** (df = 1; 5232)	27.896*** (df = 2; 5230)	23.127*** (df = 3; 5229)
<i>Note:</i>			*p**p***p<0.01

Table5– Regression Profit on Age and other variables

We've decided to use a new and tuned measure to assess company's performance. We came up with: Sales difference higher than average sales difference and called it "success".

Above you may check the results.

Once again we see that R2 coefficient is low, what may once again indicate that our models still lack explanatory power. On the first regression, and standing alone, the young variable was significant, but upon inclusion of other variables it started losing its significance.

Analyzing company's performance - part 2

	Dependent variable:		
	scorediff		
	(1)	(2)	(3)
young.x1	0.189 (1.258)	1.095 (1.316)	1.200 (1.317)
firmAge.x		0.149** (0.064)	0.120* (0.066)
ind.xManufacturing			1.686* (0.950)
Constant	-1.651*** (0.488)	-3.493*** (0.926)	-3.868*** (0.950)
Observations	5,234	5,233	5,233
R ²	0.00000	0.001	0.002
Adjusted R ²	-0.0002	0.001	0.001
Residual Std. Error	32.522 (df = 5232)	32.511 (df = 5230)	32.505 (df = 5229)
F Statistic	0.022 (df = 1; 5232)	2.749* (df = 2; 5230)	2.883** (df = 3; 5229)
Note:			*p**p***p<0.01

In order to find out if we could retrieve further insights from our data we've also considered the indicator given by:

$$\text{score} := \text{inc_bef_tax} / (\text{curr_assets} + \text{fixed_assets} - \text{curr_liab})$$

After computing the variation from 2014 to 2015 we ran a few regressions but the results were not satisfactory given that there was no significance.

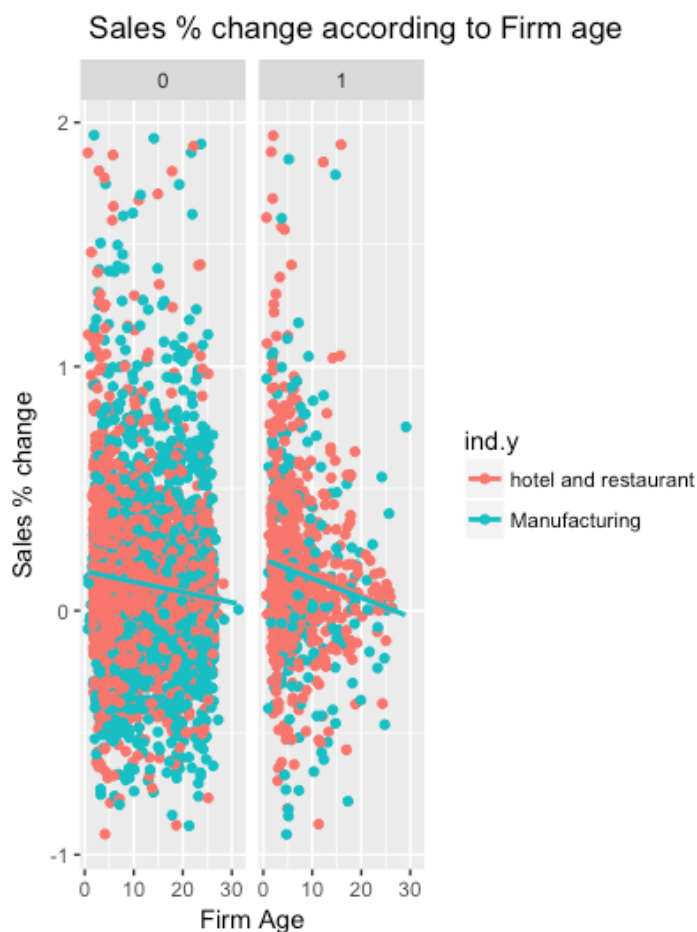
During our analysis we were able to identify that there are concentrations observations according to Firm's age and CEO's age.

Limited to our database, and as plotted below, young CEOs are concentrated on young companies, as opposite to old CEOs that usually run older companies.

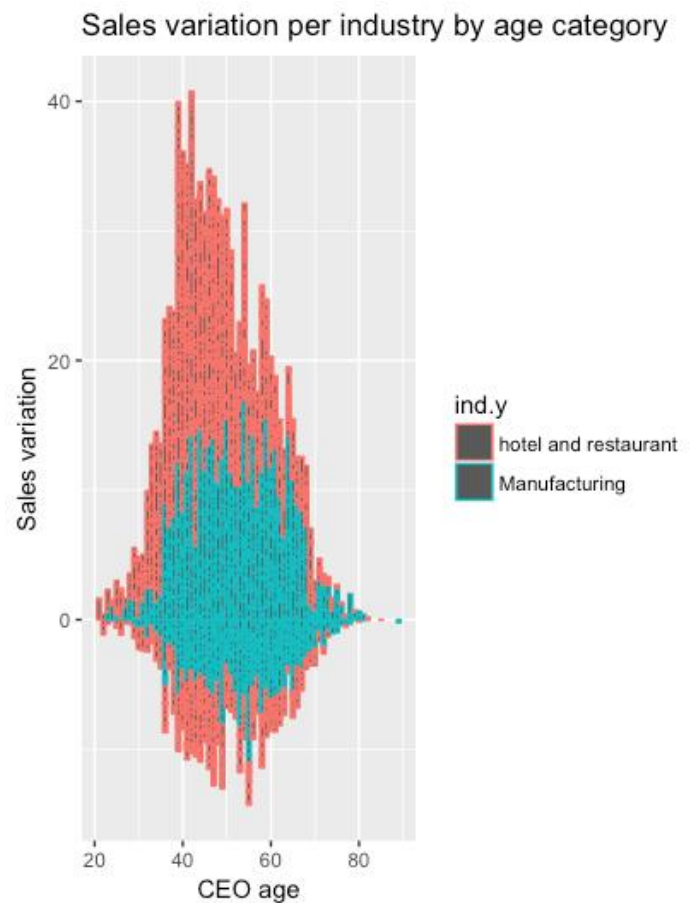
One justification for that might be that a lot of companies were created over the past few years (startups) and on average the founders may be younger.

We can also notice that manufacturing companies, generally speaking, are older than companies that operate on the Hotels and Restaurants industry. One justification for that might be the recent boom in terms of tourism and business in Hungary, which led to an increase of opportunities on this industry (more beds for tourists and business people and more restaurants for tourists, locals and foreign workers).

Plot8 – Sales % change according to firm age



Plot9 – Sales variation per industry by age category



We were asked if companies with young CEOs had better performance than companies led by old ones. Let's cut to the chase: there is no evidence of that for small and medium size corporate in this sample.

We were able to identify some correlations and relationships but with the data provided we cannot draw such a conclusion. There is no evident causality.

Even if we were asked to confirm if young CEOs performed better in Hungarian companies that operate on hotels and restaurants and manufacturing, we would have some reserves. Worse if we were to answer in general terms to later extrapolate.

The database is limited to Hungary and restricted to a few industries, which in a context like this limits the conclusions that we can come with. We need to take into consideration that sometimes even industries that operate on the same country, are constrained by different laws, management cultures, limited to licensing or even prone to a kind of management (for instance if we would consider the technological industry, the empirical knowledge tells us that most probably the majority of the CEOs – founders – are young).

Other variables should be considered, some of them beyond the CEO itself, but even if we base our analysis on the CEO (person) we should also consider her/his background, education level, experience (total and on a specific industry – for instance you may have a CEO with 45 years old that until recently was a 1st line manager vs a CEO that although with 39 years already created several companies).

Additionally we can also say that being data from one single country, culture aspects should also be considered in terms of top management election.

Concluding, the analysis would be more conclusive if there was more data from different countries, companies, agents (CEOs), and other micro and macro variables – but the world is not perfect, nor the data in it.