CONTRIBUTING INSIGHTS TO OBTAIN A SERIES B FUNDING ROUND FOR ARCUS FINANCIAL INTELLIGENCE

Submitted for the fulfilment of requirements for the degree of
Master of Science in Data Science of Goldsmiths, University of London

Department of Computing

Supervised by Dr. Ida Pu

Imanol Belausteguigoitia Ibarrola

September 9th, 2019

Abstract

The company Arcus Intelligence seeks to expand its services and grow its presence in the United States and Latin America, but for that they need to obtain financing in a Series B funding round. Arcus wants through this study to improve their understanding of the socio economic components that enable Fintech emergence in Latin America and to have powerful insights as to how the market trends are moving in recent years for B2B and B2C Fintech companies in Latin America. The aim of this project is to provide the company in question powerful insights through multiple tested hypotheses, visualizations and forecasts that provide a better understanding of the Fintech market in Latin America.

Acknowledgements

This dissertation is dedicated to my parents who have given everything for me to have an excellent professional and personal formation. Thanks also to my uncle Juan Carlos for being present in my life, listening to me and having a special connection, as well as all my family members. Thanks also to my siblings because together we have laughed and have made life a light and fun space.

The United Kingdom has a special place in my heart, it has taught me values such as respect and integration of different ways of thinking and living, that taboos must disappear and that solidarity is a wise path that we must embrace. Thanks to each of my colleagues from the Goldsmiths Data Science program especially Yino and Simone, who together spent with me countless hours in different London cafes trying to take advantage of the experience, and Esperanza, because together we managed to break many cultural barriers. Thanks to the different british and international students of the course, they all taught me many lessons that are now part of me.

This program would not be possible without the presence of Daniel Stamate, thank you for your understanding, for your receptivity and your brilliant way of being the leader of the program, you have been truly excellent. I want to thank my cousin Iñigo Rumayor for believing in me and supporting me professionally, we had great and tough moments together, of learning, and of intensely supporting each other in the professional realm.

Thanks to Ricardo, Yorgos and Javier for being present, as well as all the members of my close group of friends in Mexico. Thanks to Paramasukha, for his meditation sessions and for teaching me the importance of being connected and united with my reality. Many thanks to Marifer for the last three months of consistent support. Finally, thanks to Guadalupe Marengo for her hospitality, and amazing personality that made my stay in Hackney great.

Content

Introduction

This dissertation is designed to be highly interactive and with a great emphasis on storytelling. As such it was created so that it could be read not only by academics in the field of Data Science or Statistics, but also for public interested in fields such as Entrepreneurship and Business Development.

Some of the techniques and procedures have an experimental character and are not necessarily common in the field. The work contains personally developed techniques that have a detailed mathematical justification.The scope of the main domains of dissertation covered in the implementation phase are exploratory, explanatory and predictive.

It is key to take into account the environment in which this work was executed and to who it is directed. In this sense, the present work is going to be primarily used by a Fintech startup that does not apply Data Science or Data Analytics in their business operations. Therefore the document is designed to communicate synthetically key findings about the Fintech market in Latin America. Graphs and results are presented in a fashion that could be presented in business presentations to potential investors.

The work presented is extensive and mainly focuses on geographic region of Latin America. Overall the socioeconomic enablers of Fintech emergence in the region show favorable figures in the case of the region studied. B2B and B2C companies in Latin America show an increase in comparison to previous years and the tendency is that they will continue to increase in the following three years.

2. Background and literature review

The main objective of this section is to explain some of the most relevant concepts to understand the purpose of this dissertation. There will be reviewed and explained synthetically the concept of Fintech, the nature of startups, and the fundamentals of funding stages for startups. Finally, the context in which it Arcus is immersed today will be explored, as well as how the company will benefit from this study.

2.1 Fintech - the combination of "financial technology"

The term Fintech is used to describe new technologies that seek to operate and automate financial services. Fintech is mainly used by entrepreneurs, companies such as finance service providers and consumers to better manage their financial operations using specialized software used in mobile devices or computers.

This discipline flourished in the 21st century and was initially associated with the back-end technology used in systems of financial institutions such as banks. The concept has mutated to have a wider connotation. Fintech now includes a large number of sectors and applications such as fundraising, education, personal finance management, money transfer between peers, retail banking and many more. Fintech also includes a recent branch called crypto currencies such, where Bitcoin or Libra, the new cryptocurrency of Facebook, are architectured and used as a currency.

According to Crunchbase, there were 12,000 active Fintech in the world in 2019. The United States is the most active country in this sense, followed by India and the United Kingdom. These can be classified into several sectors, including investment management, collective financing, payments, loans, cross-border remittances, insurance, and more, depending on the specific service that a company wants to provide. Currently, there are many easy-to-use mobile applications that need no personal assistance, such as Robinhood, that do not charge commissions for transactions.

Since the last decade, the proliferation of mobile applications has been present, and particularly Fintech has experienced impressive growth as worldwide funding in Fintech continues to peak every year. One of the characteristics of Fintech's new companies is that they seek to threaten or challenge the traditional way in which banks operate, since they rely on advanced technology to provide better and faster services to different users.

In recent years Fintech has been able to capitalize on three major areas of improvement of the banking system: being very bureaucratic, opaque, and with low quality user-experience interfaces. The new technologies applied by Fintech companies, include machine learning or artificial intelligence, and applications such as the creation of customer profiles, targeted promotions are applied now. Fintech as a sector also promotes the automation of the processes carried out in banks to help their clients to use strategies such as chatbots, as well as algorithms for fraud detection by tracking transaction history of individuals.

2.1.1 Fintech business making levels

There are three types of business models in Fintech. For that, it is important to know what their initials mean:

C = customer
B = company

You can make three basic business combinations with these two letters: (1) B2B is when the service or product is offered in an interaction between two companies, the most common companies are processors or specialized software providers for financial service providers,[1] (2) C2C, one of the most emblematic cases is Venmo, where two customers to execute a transaction, and (3) B2C, which is when the service is carried out from a business to a client, a very common example are commercial banks, as Barclays.[2]

Arcus, the company that will be benefited from this work focuses on B2B, which signifies business-to-business. In this business model, the company in question centers on providing products and services to other companies. In this sense, the company becomes a supportive force through their services and assist their clients in their internal processes.

2.2 Defining startups: potential is the key

A startup is a young company that offers a product or service that the market does not currently offer or that it is covered in a basic way. There is no reliable data or perfectly clear rules that define a startup in terms of profits or expenses, valuation, employment and the parameters vary considerably according to the sector. It is also difficult to know how old the company has to be, according to Accelerator Y Combinator leader Paul Graham, a company with more than five years that can continue to be a startup

Approximately three years after its creation, startups stop being startups, but that is only a trend. This depends on several situations, a startup can cease to be if it is acquired by a large company, or has more than $10 million[3] in profits, more than five people on the board of directors, or more than 80 employees. While there is a gray area to define what a startup company is a fundamental variable is the ability of the company to approach a consolidated cash flow and high positive revenue numbers.

Although this is generally the case, startups do not always have to be technology companies. For a company to be a startup, it must be accredited as a new option with high expectations of

[1] Arcus, the company explore throughout this dissertation does B2B business.
[2] When it comes to consumers, the younger the individual, the easier it is to have an idea of what financial technology because there is a tendency for them to be related to mobile application and new technologies. In this regard, it should be noted that Fintech technology is more customer oriented towards the millennial market due to its great success and the expected potential of this strategy.
[3] The currency used for this dissertation is USD

growth and expansion in the future. That's why it is associated with technology and highly scalable software while restaurants or similar companies in early stages don't fit these requirements.

2.3 Funding stages

All big companies start with an idea that was executely not only administratively but also economically. In most cases, companies finance their first months of activity with the help of relatives, friends or the founder's own financial resources. If the company proves to have a successful product or service, its client will probably grow and the company will begin to develop its market share and ambitions. In extremely successful cases, the company acquires great value over time and extends to different countries to include more products, managers and investors, and eventually is made public in an initial public offering (IPO).

When funding Series A, B or C are mentioned, they refer to the successive stages of a company's growth through external investment. Financing offers investors the opportunity to invest money in exchange for shares or to keep a part of the company. Most of the successful startups at Fintech have raised funds through external funds provided by entities such as venture capital funds or investment banks.

It is very common for new companies to get involved in so called seed funding or angel funding in their early steps. The resources that these companies can acquire are used to carry out their first operations and spend in different strategic areas to develop the business and make it more profitable. Then, if the company in question requires more financing, they can request more funding in Series A, B or C, which will be explained in the following paragraphs.

The path that each company takes will depend on the size of the market it seeks to cover the operating costs and the speed with which the sector to which it belongs moves. Some companies take a few weeks to find their first investors (especially in the field of technology) and soon seek to make the company public, while others, with minor ambitions and slow growth, will receive little money from their relatives and end external financing.

Understanding the distinctions of each of them allows us to put into perspective the process of a hypothetical business, its size in economic terms and its intentions (in the most extreme case of growth and expansion, it will become an initial audience). Seed and pre-seed funding, Series A, B and C are the stages in which entrepreneurs spend injecting resources into their businesses.

2.3.1 Funding components

In order to have a basic and precise understanding of funding, it is necessary to discuss three main concepts: the recipients, the investors and the evaluation of the company.

The recipients are part of the transaction who look for resources to operate a business idea or an established business, in search of growth or profitability. As the business begins to mature, it tends to advance in the financing rounds, as mentioned above. The interaction between the recipients and the potential investors is due to the fact that the investors want the business to prosper and believe in the causes of this business and their business idea, although they also seek to generate economic returns through the injection of money into the social capital of the company in question.

Before any round of financing, the investor's analysts make an assessment of the company in question. Valuations depend on a number of factors, including management, performance, market size and company risk. One of the keys to distinguish financing series refers to the valuation of companies (the higher the market value, the more generally the series are advanced), as well as the company's maturity and growth prospects.

2.3.2 Pre-seed and seed funding: the initial push

The first stage of financing a new company is so early that it is often not considered in the company's records. This stage is known as "pre-seed" funding, it generally refers to the period during which the founders of the company are the only ones who work for the company. The most common scenario is that they finance the business, but it is not uncommon that they receive financing from close relations. Depending on the nature of the business, the initial investment will be different, this financing step can be instantaneous or long. In most cases, those who currently invest in the company do not invest in company shares.

Seed funding is the first formal step in a company’s funding record. In general, it represents the first external economic resources that a company collects. Many companies do not go beyond seed funding and do not move towards a Series A or B. To be able to receive external financing, entrepreneurs must present a convincing business plan and a successful strategy to their potential investors.

One of the most common investors in this stage are called *angel investors*, who like to make riskier bets when they bet on commercial projects that have few statistics, but that could experience strong growth in the future. The amounts collected during a seed stage vary from one case to another, although it is common to be in the range of $ 10,000 to $ 2 million, depending largely on the first resources to operate the company. Most companies in the United States have net value between $3 and $6 million after obtaining seed funding.

2.3.3 Series A: the company running now

Once the company can operate and provide products and services for a specific market, maintain a customer base, consistent results figures and indicators. If the company seeks to increase their performance, their managers can look for a Seria A to optimize its services and think to expand its market to a wider audience or to develop new products. Series A investors are usually strong venture capital firms. Sequoia Capital, Khosla Ventures or Accel Partners are some of the most famous in the market.

In a Series A, between $2 and $5 million are generally collected, but these figures are constantly increasing due to the recent proliferation of companies that are growing exponentially in short periods of time. These types of companies are called "unicorns" in the entrepreneur realm. These are strange cases that are generally backed by venture capital for more than $1 billion and remain private. Most of these companies have a high-tech component and software and do not necessarily have a large number of employees compared to their profits. Some of the biggest unicorns that operate today are AirBnb or Uber.

At this point, it is common for investors not only to support financial resources, but also to participate in the management of the company, not as employees but as supervisors of their investments. Some of the investors in this step can also be *angel investors*, but it is not very common. It is increasingly common for companies at this stage to use crowdfunding to gather the resources they want. Entrepreneurs who achieve a successful A Series should consider their business as less than half of them spend initial funds in Series A

2.3.4 Series B: looking for expansion

The particular reason for performing a Series B is to extend the activity beyond a stage of development. Investors in this stage are helping start-up to expand their market. Many companies that move from the Series A to Series B have shown that they have enough stability and a solid base of clients to convince investors that, with a greater injection of resources, even more business.

In this sense, the company in question has already validated a market and a business idea and successfully exploits part of the market to which it belongs. For this step, the company generally seeks to invest in business development, marketing, technology or the acquisition of highly qualified employees and managers. A common figure for raising capital at this stage is $10 million. Companies seeking a B Series have higher average valuations than those in the development phase with market values that usually range between $30 and 60 million.

The nature of Series B investors is very similar to Series A, although it is common for companies to be financed with mature growth funds because the risk associated with this investment is generally less because they have a successful record.

2.3.5 Series C: going public or acquiring other companies

The companies that enter a Series C are already established in the market and their valuations are well above average. This type of companies are looking for funds to develop new products, to create new business sectors for a transnational market in many cases, although they usually go to a Series C to acquire new businesses or go public.

At this stage, investments become less risky and more investors tend to participate. In a Series C, investing parties similar to Series A and B generally such as investment banks, hedge funds, private equity firms appear. Investors who choose this bet often have many resources, unlike small size investors that can contribute to the initial stages.

In most cases, the company will stop seeking external financing, although some companies seek to continue with the D or E Series. These types of companies already operate in several countries and have a valuation of several hundred millions of dollars, but are looking to expand globally. There is also the case of companies that make a D or E round, not necessarily for a positive reason, some companies are turning to the D or E Series for not being able to achieve their expansion objectives in a C Series and are looking for a second attempt in subsequent cycles.

2.4 Haddad's Global Fintech Market research paper

This work is inspired in the paper "The Emergence of the Global Fintech Market: Economic and Technological Determinants." from author Christian Haddad. In his work Haddad seeks to statistically prove which technological and macroeconomic components have an influence in the emergence of new Fintech companies in a Worldwide scale. The study takes as a sample 55 total countries and takes extracts variables from different public sources such within a time period from 2006 to 2014 and covering 5588 fintechs.

As such, Haddad does his study at a considered worldwide scale and uses regression analysis to reject or prove a set of hypotheses related Fintech startup formation. The papers centers on the ability of policymakers to shape the emergence of the Fintech markets in the geographic region they belong to and how does regulation and socioeconomic components enable and shape the macro trends of the sector.

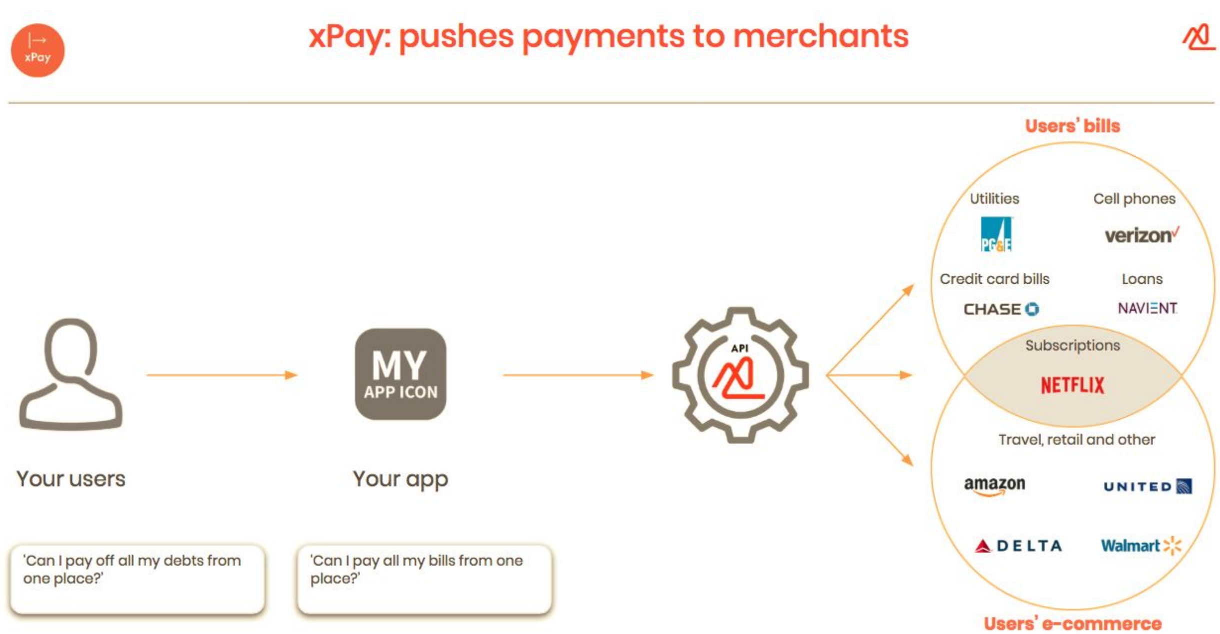
2.5 Arcus in the Fintech landscape

In 2013, Arcus (called then regalii) was founded by Edrizio De la Cruz, Iñigo Rumayor Belausteguigoitia and Juan Maldonado in the Y-Combinator of Silicon Valley. Initially, the company offered cross-border bill payments focused on the Latin American immigrant market in the United States. Subsequently, Arcus evolved in 2016 to focus on the payment of invoices in the domestic market.

To date, Arcus has raised a total of $ 18.9 million in Series A and will seek to raise $ 20 million in a Series B by the end of 2019 and early 2020. Some of its investors are IGNIA, Andreessen Horowitz, Winklevoss Capital Management, HOF Capital and Y Combinator.

Arcus develops software for companies and, therefore, is considered considered a B2B company. Currently, it mainly manages three products: xData, xPay and xChange[4]. The company is headquartered in New York and Mexico and focuses on programmatic payments and card updates, as well as data mining, all from any biller.

Figure 1: Arcus products: xPay

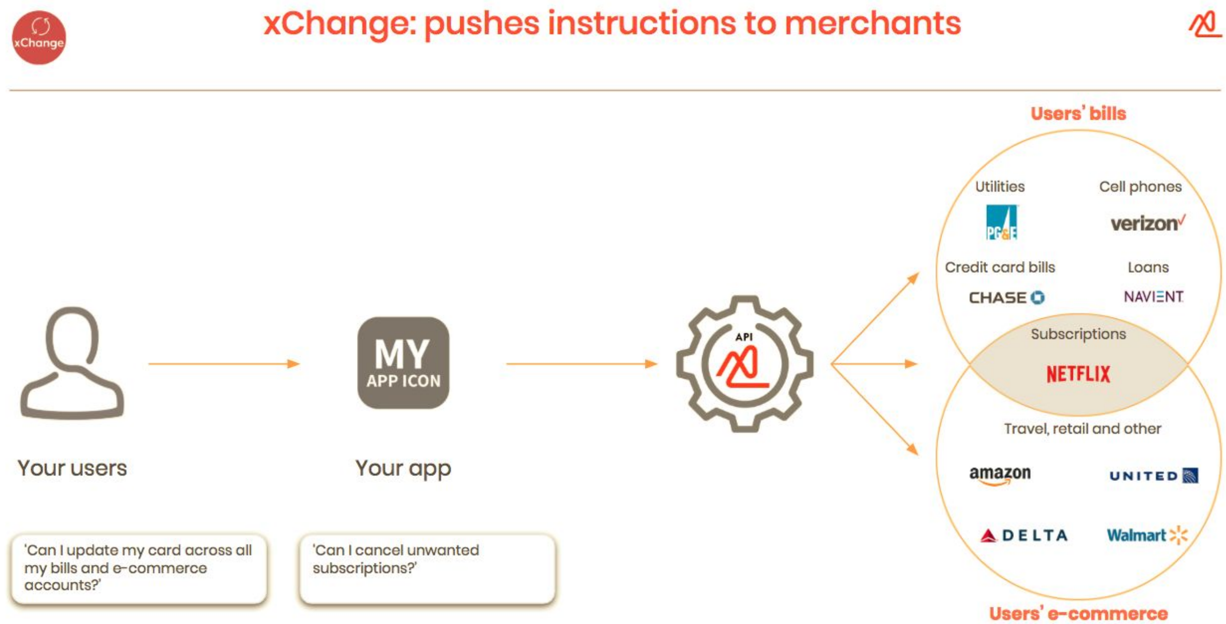


extracted from Arcus company presentations

Figure 2: Arcus products: xChange

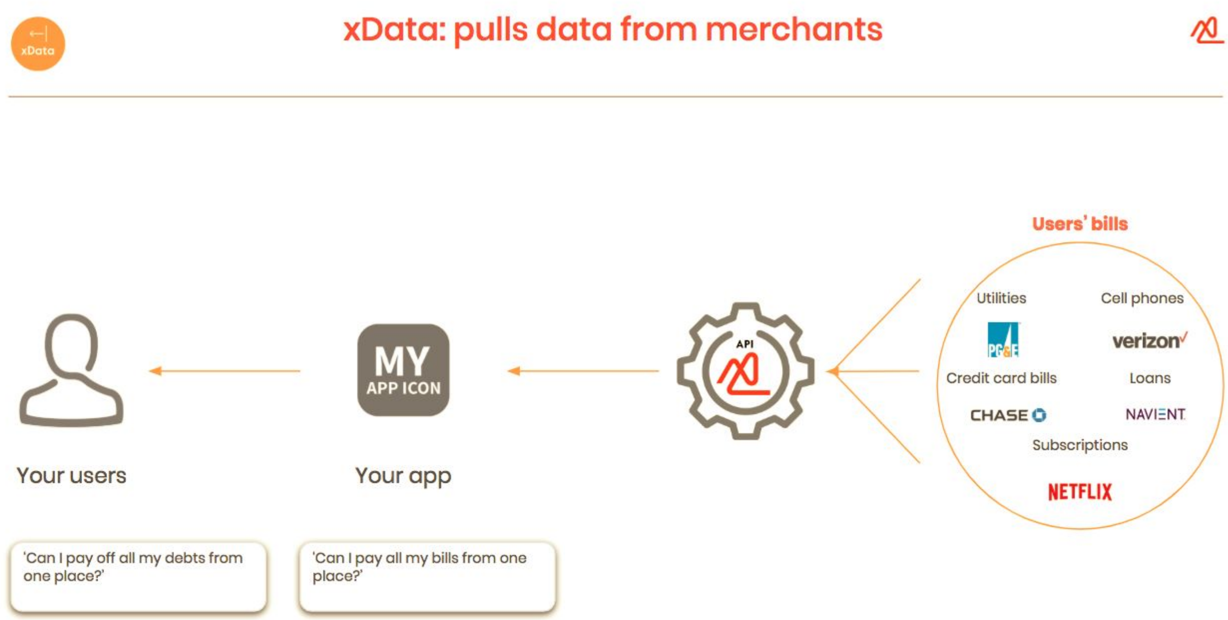
---

[4] xData enables bank users to track payment data and get real time customer insights to help drive their finances. xPay is a procedure to pay for everything, including bills, credit card debt, utilities, student loans, rent, and much more. xChange automatically updates card-on-file information, ensuring uninterrupted service and payments for cardholders. For more information visit **https://www.arcusfi.com/products/**



extracted from Arcus company presentations

Figure 3: Arcus products: xData



extracted from Arcus company presentations

Today, Arcus has the slogan "The new infrastructure for recurring payments" and offers services large international banks such as Chase and Bank of America. Arcus has been active in the market for almost six years, but it is considered a startup, since it offers a product or service that is not currently offered in the market or that is underdeveloped and is considered to have very high growth prospects.

To archive Arcus' main goal for 2020, which is to complete a $20M Series B for 2019, the focus is to culminate the long sales cycle process and launch the largest enterprises in the industry, including Bank of America ("BofA"), J.P Morgan Chase ("JPMC"), Barclays, Marqeta, Acorns in the U.S., and Citi Banamex, BBVA Bancomer, BanRegio, Scotiabank, Rappi, Uala in LatAm.

The Arcus focus for the following months, until 2020 will be to continue investing in the necessary infrastructure to launch and support the largest FSPs in the market. The company will also focus on expanding merchant partnerships.

The purpose of this study is to provide the company with a better understanding of the Fintech landscape in Latin America through an exploratory analysis of this topic, hypothesis testing and times series analysis. With the assistance of complex algorithms and content studied in the MS of Data Science of Goldsmiths, University of London this work will portray powerful insights that will assist Arcus to better understand the Fintech market in Latin America, the recent trends it experiences, and the particular socioeconomic drivers that enable the emergence of the sector in question. This project aligns to the company's objective to informedly address raise funding for a Series B in 2020.

3. Design

This work combines exploratory analysis through summary statistics tables and data visualizations and explanatory analysis by hypothesis testing for multiple geographic region although it is focused in Latin America.[5] The dissertation finishes with a time series analysis to forecast Fintech emergence in the next three years in Latin America.[6]

3.1 Workflow

This implementation of this dissertation consists of three parts:

part I: International case

[5] Specifically the sample frequently used contains all Southamerican countries and Mexico.
[6] I have a personal motivation to explore for the first time time series.

The main objective of this section is to explore the significance of six hypotheses explored by Christian Haddad in a different time period (2009-2009) for a different sample of countries (17) with a self developed technique (weighted variable testing Pearson correlation one tailed test).

part II: Latin American case
Part I will be contrasted with a sample of Latin American countries. In particular the significance level of the hypotheses will be compared against the international case (part I) and the overall shift of components in recent years will be visualized and tested.

part III: Fintech B2B-B2C market trend analysis
A set of new hypotheses personally assigned will be tested with a Student's T test to gain powerful insights of Fintech market trends of recent years in Latin America. Times series analysis will be applied to forecast new Fintech company emergence in the next three years.

**The results of the hypotheses and the contrast among the regions studied will not be explored in the work presented. However they may be explored in future works or discussed in the viva session.*

4. Technical Applications

Programming Language: Python
Programming Environment: Google Colab
Data Visualization: Matplotlib*[7] , seaborn* , Google Sheets graphics.
Data Engineering: Google sheets, pandas* , fancyimpute* , numpy*, sklearn*
Data extraction: Crunchbase Pro, World Bank, and World Economic Forum.
Hypothesis testing: scipy.stats*
Machine Learning: sklearn*, statsmodels*
Data Science techniques and domains: imputation with predictive algorithms, data pre processing, hypothesis testing, time series analysis, variable architecture, data visualization, machine learning, exploratory analysis, storytelling.

[7] * *package from Python*

5. Implementation

Part I.

Socioeconomic component analysis International case

Dataset

“Mastersheet_part1.csv”: shape 21 x 171

	Countries fintechs	Country	Year	1A	1B	1C	2A	2B	2C
0	4.0	Argentina	2018.0	2.447560	2.805120	4.061309	3.430802	4.195683	3.930331
1	1.0	Bangladesh	2018.0	2.578552	3.683762	4.117619	3.058403	3.355734	3.716894
2	32.0	Brazil	2018.0	2.459301	3.302393	4.566283	3.238514	4.296651	4.089989
3	30.0	Canada	2018.0	3.689212	4.582466	5.105235	4.659581	4.638758	4.652790
4	1.0	Chile	2018.0	3.504282	4.101409	3.883189	3.885442	3.263582	3.997062

Sources

Crunchbase *, World Economic Forum “World Competitiveness report” **, World Bank***

Target variable

Countries Fintechs* : the number of founded Fintechs in a particular year for a specific country.

Independent variables used

Venture Capital availability** → renamed as “1A”
Availability of financial services** → renamed as “1B”
Affordability of financial services**→ renamed as “1C”
Quality of education** → renamed as “2A”
Internet access in schools** → renamed as “2B”
Availability of latest technologies** → renamed as “2C”
FDI on technology** → renamed as “2D”
Individuals using internet** → renamed as “2E”
Availability of financial services** → renamed as “3A”
Affordability of financial services** → renamed as “3B”
Access to loans** → renamed as “3C”
Soundness of financial sector** → renamed as “3D”
Mobile broadband subscriptions** → renamed as “4A”
Mobile telephone subscriptions ratio** → renamed as “4B”
Domestic market size index** → renamed as “5A”
Foreign market size index** → renamed as “5B”
Unemployment rate*** → renamed as “6A”

Time period

2009 - 2019

Countries

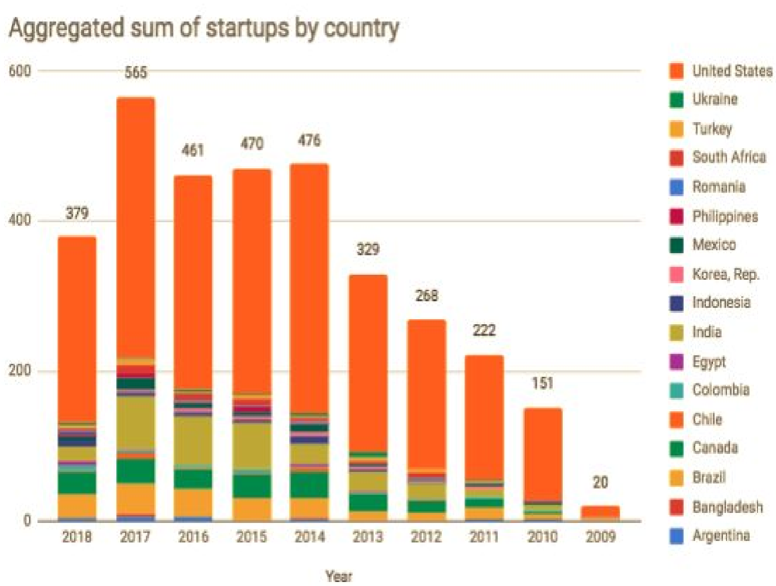
Two criteria is to get a (1) balanced set of countries in terms of economic power and development including 5 highest GDP economies in LatAm, (2) Information completeness.

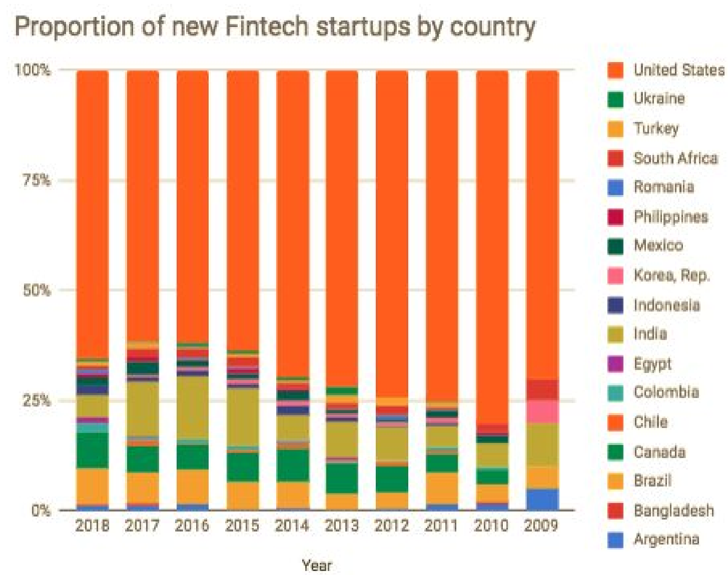
5 highest GDP in Latin America: Brazil, Mexico, Colombia, Chile, Argentina
3 High income United States, Rep. of Korea, Canada
3 middle high income Turkey, South Africa, Romania
5 lower middle emerging emerging: India, Egypt, Indonesia, Bangladesh, Philippines
1 lower middle income: Ukraine

<u>Completeness</u>
The following columns contained missing values: ["1B", "1C", "2B", "2D", "2E", "3A","3B","3C","4A","5B","5C"]
this columns not filled in the extracted information: **deterministic linear model** was used to estimate the missing values.

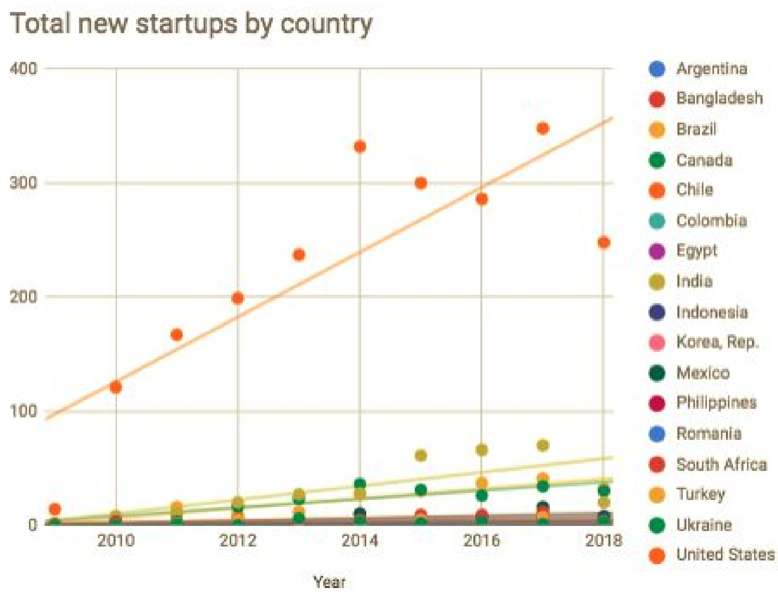
Exploratory Analysis part I

The following graphs explore new Fintech startups emergence in the sample. The first graph shows the aggregate sum of new companies, being 2018 the year with most new startups (565) and 2009 year with least (20). The second visual illustrates that the United States concentrates more than Fintechs followed the sample followed by India in every .

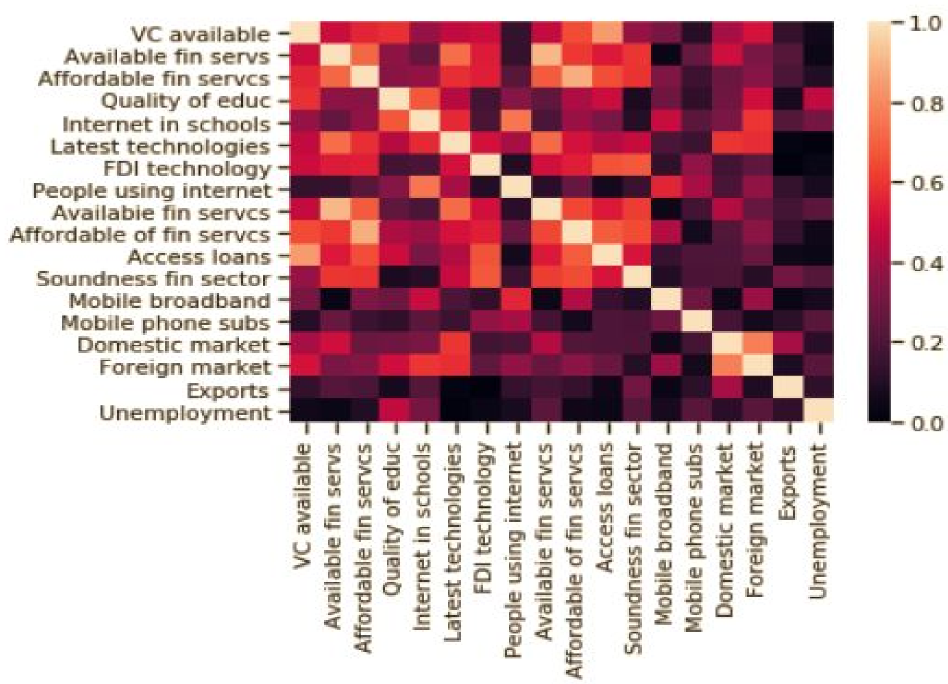




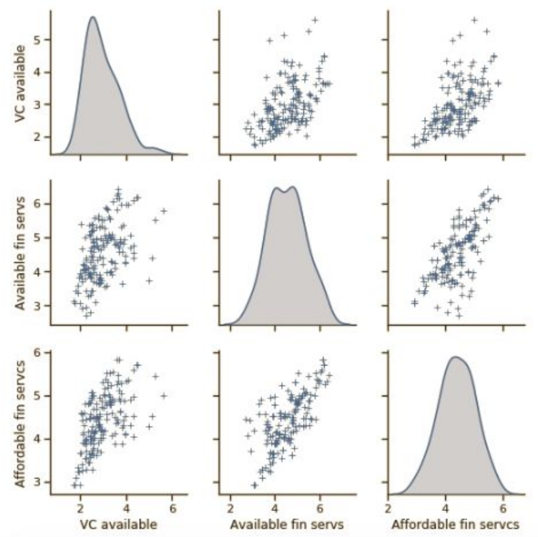
There is a clear growing tendency of emergence of new companies illustrated by the tendency lines.

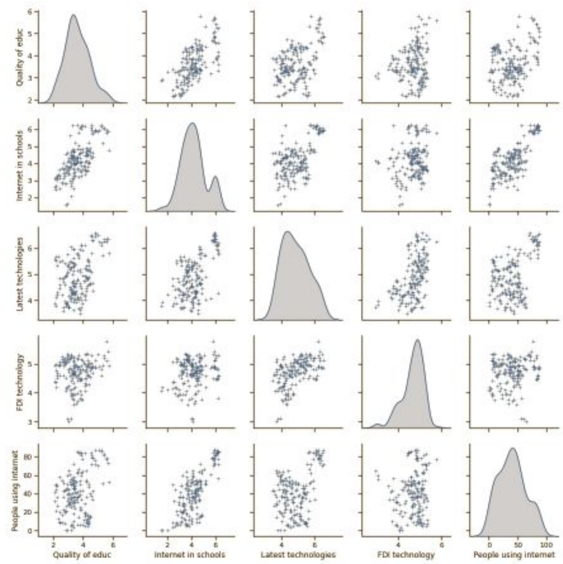


The next graph shows correlation in a heatmap fashion. Lighter colors demonstrate high correlation in absolute number, while dark colours show the opposite.

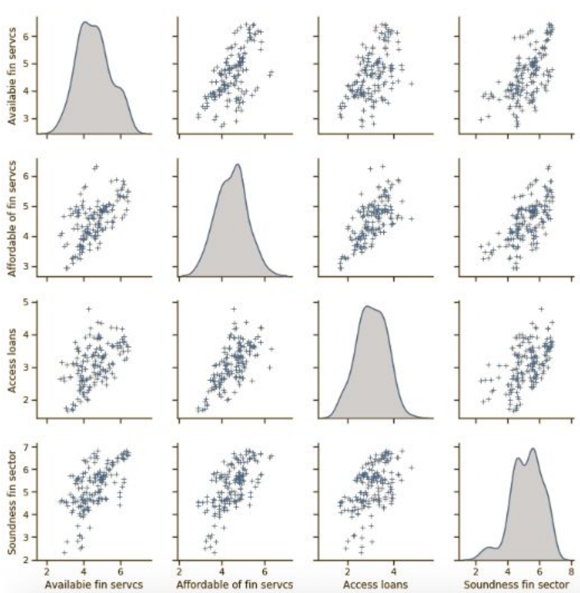


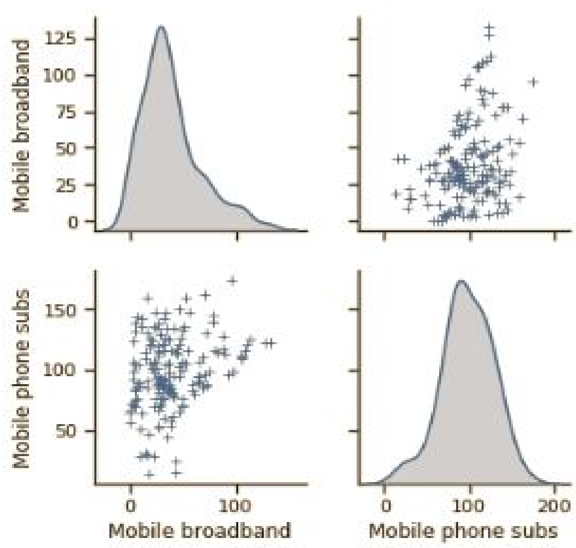
The graph on the left shows contain distribution and pairplots for the following capital development related variables in the following order: Venture capital availability, Availability of financial services, Affordability of financial services. The second one contains the same information for technology available variables: Quality of education, Internet access in schools, Availability of latest technologies, FDI on technology, and Individuals using internet.



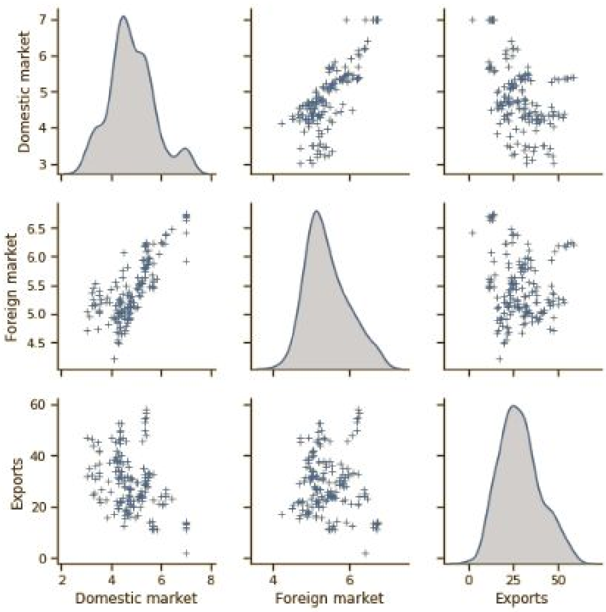


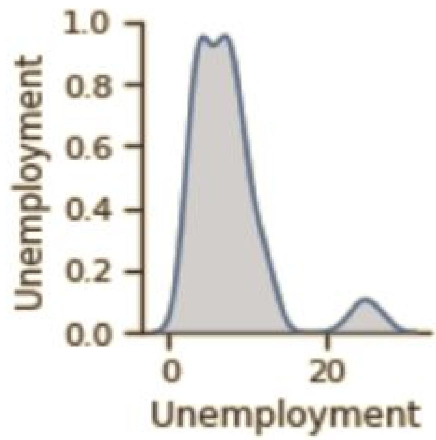
Below is the normality and pairplot relation among financial soundness variables (5) and mobile phone use variables (2).



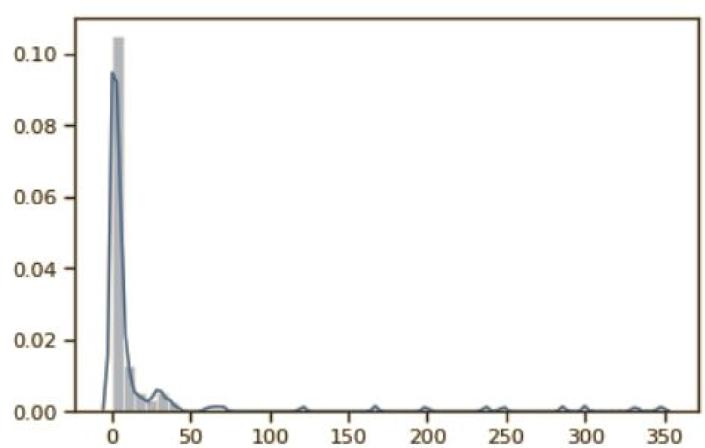


The next visuals is the normality and pairplot relation among financial market size variables (3) and the unemployment variable, which was extracted from the World Bank data.





The next visual illustrates the target variable which contains the total new Fintech startups in periods of one year from 2009 to 2018. It demonstrates that the data does have a normal distribution since most data points concentrate from zero to 50, but there is a major outlier, which is the United States which consistently has more than 100 new Fintech startups per year.

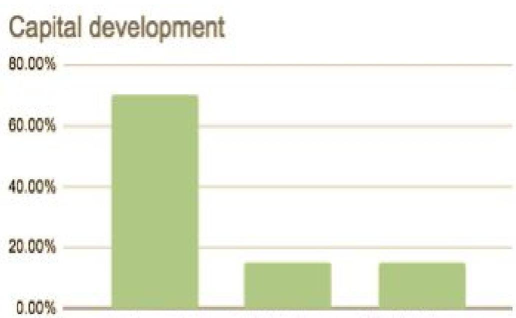


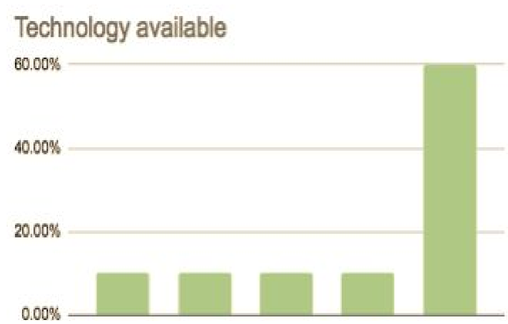
Weighted variables

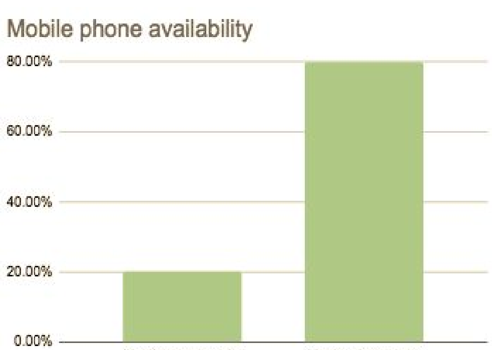
As seen in the previous section, each variable used for this part of the dissertation belongs to a category which can be, capital, technology, financial, mobile telephone, market size, or unemployment. Each category has a called *master variable* that is composed of the weighted values of all the variables that are contained in that category and they all sum 100% of the variable. The master variable contains a percentage of the contingent variables like follows:

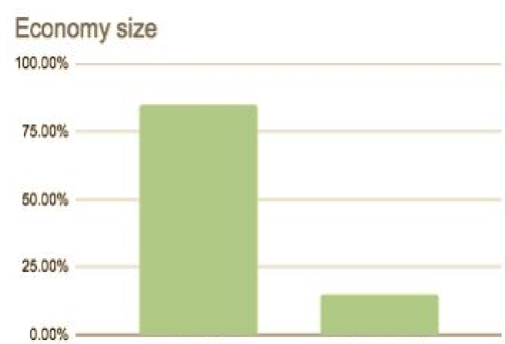
Master capital = (1A*0.7)+(1B*0.15)+(1C*0.15)
Master technology = (2A*0.1)+(2B*0.1)+(2C*0.6)+(2D*0.1)+(2E*0.1)
Master financial = (3A*0.05)+(3B*0.05)+(3C*0.05)+(3D*0.85)
Master mobile phone = (4A*0.2)+(4A*0.8)
Master economy size = (5A*0.9)+(5B*0.1)
Master unemployment = 6A

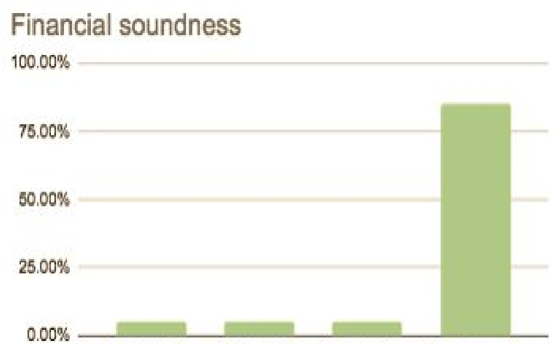
This entails that each value contained in a master variable is the sum of the weighted values of the variables contained in that category for each row. The next graphs demonstrated the weighted value of variables for each master variable.

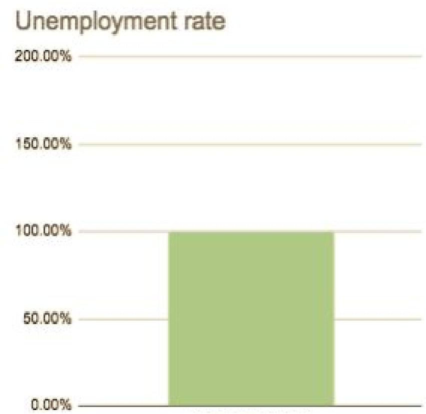












Data Preprocessing

Data standardization was performed for master variables with preprocessing module from sklearn Python package.

The command was the following:

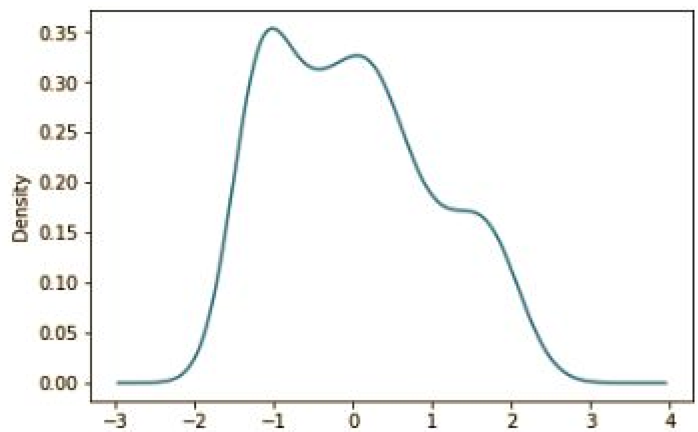
- *df_processed = preprocessing.StandardScaler().fit_transform(df_weights)*[8]

Below are the summary statistics of the center and scaled dataset where mean of values is 0 with a marginal difference and the standard deviation is 1 with a miniscule deviation.

	Master_Target	Master_one	Master_two	Master_three	Master_four	Master_five	Master_six
count	1.700000e+02	1.700000e+02	1.700000e+02	1.700000e+02	1.700000e+02	1.700000e+02	1.700000e+02
mean	1.345329e-16	2.533921e-16	1.253899e-16	5.498869e-16	3.608225e-16	4.649875e-16	9.992007e-17
std	1.002954e+00	1.002954e+00	1.002954e+00	1.002954e+00	1.002954e+00	1.002954e+00	1.002954e+00
min	-1.239331e+00	-2.006937e+00	-1.921928e+00	-2.969766e+00	-2.730663e+00	-2.021888e+00	-1.068088e+00
25%	-9.420275e-01	-7.363685e-01	-6.683470e-01	-6.644969e-01	-5.917480e-01	-6.015174e-01	-6.728093e-01
50%	-9.102850e-02	-1.492681e-01	-1.856960e-01	1.960577e-01	-4.653714e-02	-1.801731e-01	-1.612632e-01
75%	7.129175e-01	6.913757e-01	3.184062e-01	6.281572e-01	7.117465e-01	6.796194e-01	2.574732e-01
max	2.219003e+00	3.270443e+00	2.311652e+00	1.700850e+00	2.795885e+00	2.388033e+00	3.799977e+00

However, as seen in the following visual. The variables not necessarily have a normal distribution. Below is the distribution of the target master variables which is center and scaled:

[8] *df_weights* contains the unscaled values of a master variable dataset.



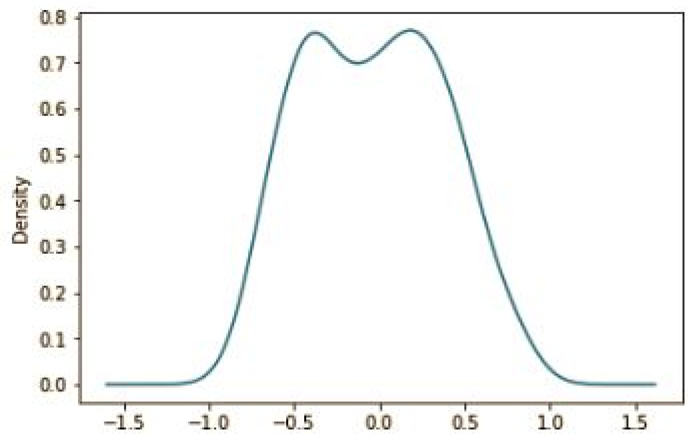
To assess lack of normality in the distribution, the *Normalizer* command from the preprocessing module of sklearn was used as below:

- *normalizer = preprocessing.Normalizer().fit(df_processed)*

Now the remaining dataset has the following summary statistics, where the mean, the range of values, and deviation, differ from the previous table.

	Master_Target	Master_one	Master_two	Master_three	Master_four	Master_five	Master_six
count	170.000000	170.000000	170.000000	170.000000	170.000000	170.000000	170.000000
mean	-0.035363	-0.039938	-0.027425	0.018928	-0.020067	-0.024902	-0.023457
std	0.405696	0.373298	0.353210	0.390303	0.385002	0.383123	0.352580
min	-0.796123	-0.784726	-0.653000	-0.854506	-0.814434	-0.855128	-0.725385
25%	-0.394860	-0.349482	-0.302073	-0.281068	-0.323934	-0.277682	-0.295066
50%	-0.045881	-0.083153	-0.098409	0.097145	-0.017768	-0.077127	-0.070114
75%	0.275142	0.311096	0.178392	0.353097	0.307789	0.288469	0.109393
max	0.811511	0.846386	0.724413	0.730769	0.850500	0.693979	0.924303

Note that the distribution of the target variable now appears similar to a normal distribution.



Hypotheses

The following are considered important hypothesis to determine their result in order to gain valuable insights of socioeconomic factors the emergence of new Fintechs in our sample.

- **H1:** How capitalized an economy is, positively affects the emergence of new Fintech companies.
- **H2:** Available technology in countries foster the emergence of new Fintech companies.
- **H3:** Sound financial systems enable Fintech company emergence.
- **H4:** The mobile subscription total number is a determinant for the emergence of new Fintech companies.
- **H5:** A large market size of an economy relates positively to the emergence of new Fintech companies.
- **H6:** Higher levels of unemployment fosters emergence of new Fintech companies.

The method applied for all hypotheses is one tailed test, with Pearson's Correlation Coefficient to test whether two variables have a linear relationship. In this sense, the master target variable will be tested against the 6 master variables to determine the result of each hypothesis.

As seen explored in previous section the assumptions to perform this statistical test are fulfilled, which are the following:

- Observations in each variable are independent.
- Observations in each variable are normally distributed.
- Observations in each variable have the same variance.

Below is the resulting correlation between master variables.

	Master_Target	Master_one	Master_two	Master_three	Master_four	Master_five	Master_six
Master_Target	1.000000	0.222047	0.369964	0.220902	0.243976	0.345109	-0.069661
Master_one	0.222047	1.000000	0.459184	0.385297	-0.084341	0.269466	-0.121491
Master_two	0.369964	0.459184	1.000000	0.422635	0.082841	0.465691	-0.115146
Master_three	0.220902	0.385297	0.422635	1.000000	-0.150941	0.134943	0.163930
Master_four	0.243976	-0.084341	0.082841	-0.150941	1.000000	-0.224944	0.127224
Master_five	0.345109	0.269466	0.465691	0.134943	-0.224944	1.000000	-0.073717
Master_six	-0.069661	-0.121491	-0.115146	0.163930	0.127224	-0.073717	1.000000

The hypothesis interpretation is as follows:

- **H0 or null hypothesis:** the two variables are independent.
- **H1 or alternative hypothesis:** there exists a linear dependency between the variables.

Results

H1: How **capitalized an economy** is, positively affects the emergence of new Fintech companies.
p value: `0.003611959034829`
ACCEPT ALTERNATIVE HYPOTHESIS

H2: **Available technology** in countries foster the emergence of new Fintech companies.
p value: `6.853359914525474e-07`
ACCEPT ALTERNATIVE HYPOTHESIS

H3: **Sound financial systems** enable Fintech company emergence.
p value: `0.003793977886372`
ACCEPT ALTERNATIVE HYPOTHESIS

H4: Mobile subscriptions is a determinant for the emergence of new Fintech companies.
p value: `0.001345110752896`
ACCEPT ALTERNATIVE HYPOTHESIS

H5: A large **market size** relates positively to the emergence of new Fintech companies.
p value: `4.048732405146512e-06`
ACCEPT ALTERNATIVE HYPOTHESIS

H6: Higher levels of **unemployment** fosters emergence of new Fintech companies.

p value: `0.366701975270771`
REJECT ALTERNATIVE HYPOTHESIS

Accepted alternative hypothesis ordered by level of significance:

technology (Hypothesis 2) > market (Hypothesis 5) > mobile phone (Hypothesis 4) > capital (Hypothesis 1) > financial sector (Hypothesis 3)

Part II.
Latin American Case

Dataset
"Mastersheet_part2.csv": shape 21 x 131

	Countries_fintechs	CB_Ranking	Country	Year	oneA	oneB	oneC	twoA
0	4	229	Argentina	2018	2.447560	2.805120	4.061309	3.430802
1	32	300	Brazil	2018	2.459301	3.302393	4.566283	3.238514
2	1	424	Chile	2018	3.504282	4.101409	3.883189	3.885442
3	7	175	Colombia	2018	2.886229	3.680098	3.975811	3.878223
4	6	176	Mexico	2018	3.152481	3.589537	4.556806	3.626325
5	1	316	Bolivia	2018	2.150270	3.869697	4.496598	3.345480
6	2	153	Ecuador	2018	2.353610	3.398379	4.410720	3.717029

Target variables

For this part, the target variable will be a combination of two variables: Countries Fintechs and CB_Ranking.

Countries Fintechs : the number of founded Fintechs in a particular year for a specific country.
CB_Ranking: company ranking assessed by CrunchBase.

Independent variables

Independent variables are the same as in part I.

Time period
2009 - 2019

Countries
13

The criteria is the geographic region where the company in question wants to expand which is South America and Mexico. The countries are:

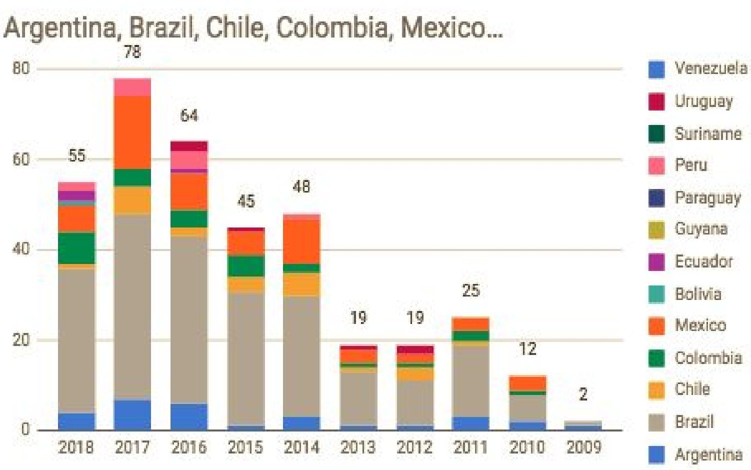
- Mexico, Brazil, Argentina, Colombia, Chile, Peru, Paraguay, Uruguay, Bolivia, Venezuela, Ecuador, Suriname and Guyana.

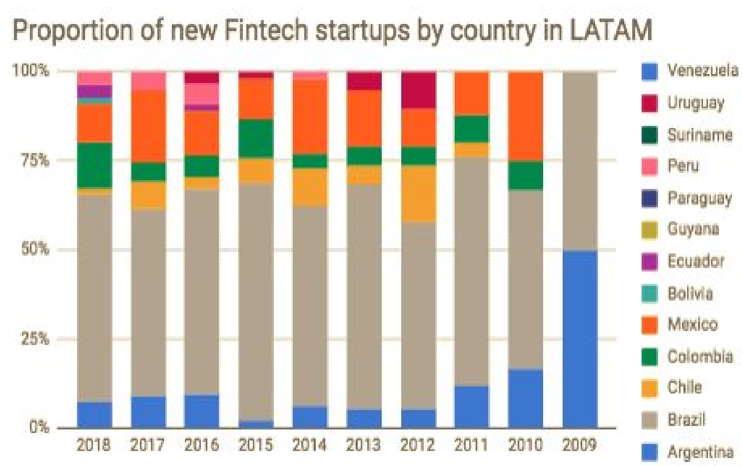
Completeness

The following columns contained missing values: : ["1B", "1C", "2B", "2D", "2E", "3A","3B","3C","4A","5B","5C"] this were not filled in the extracted information. **Iterative imputation by fancy impute* package from Python** was used to estimate the missing values.

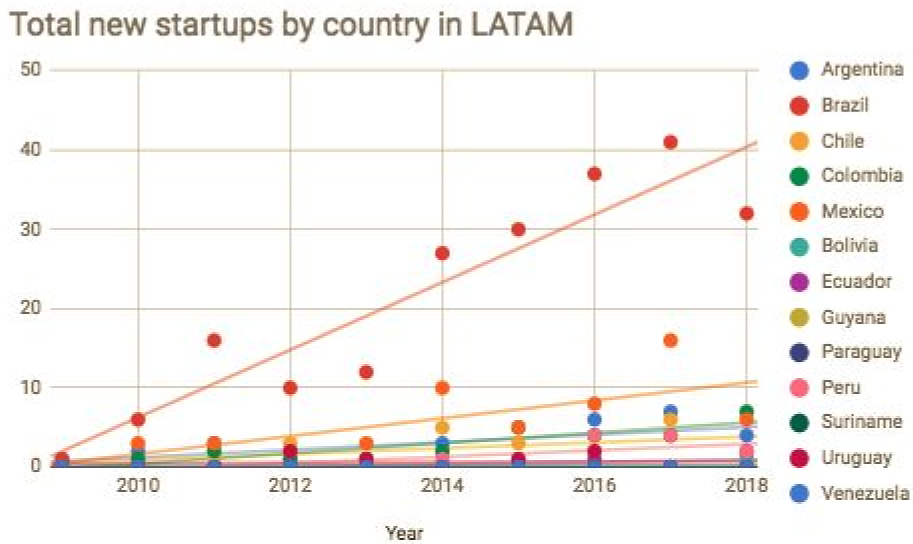
Exploratory Analysis part II

In this sample, Brazil dominates the sector and as seen in the left picture the first 5 years of the sample (2009-2013) average less number of new Fintechs than the second half (2014-2018) and 2017 is the year with more new Fintech companies in the region. The second visual in order complements the fact that Brazil has the largest share of the market, however Mexico and Argentina appear to be in the second and third position. Other countries have a represent a minor part of the Fintech companies founded in the region in that time period.

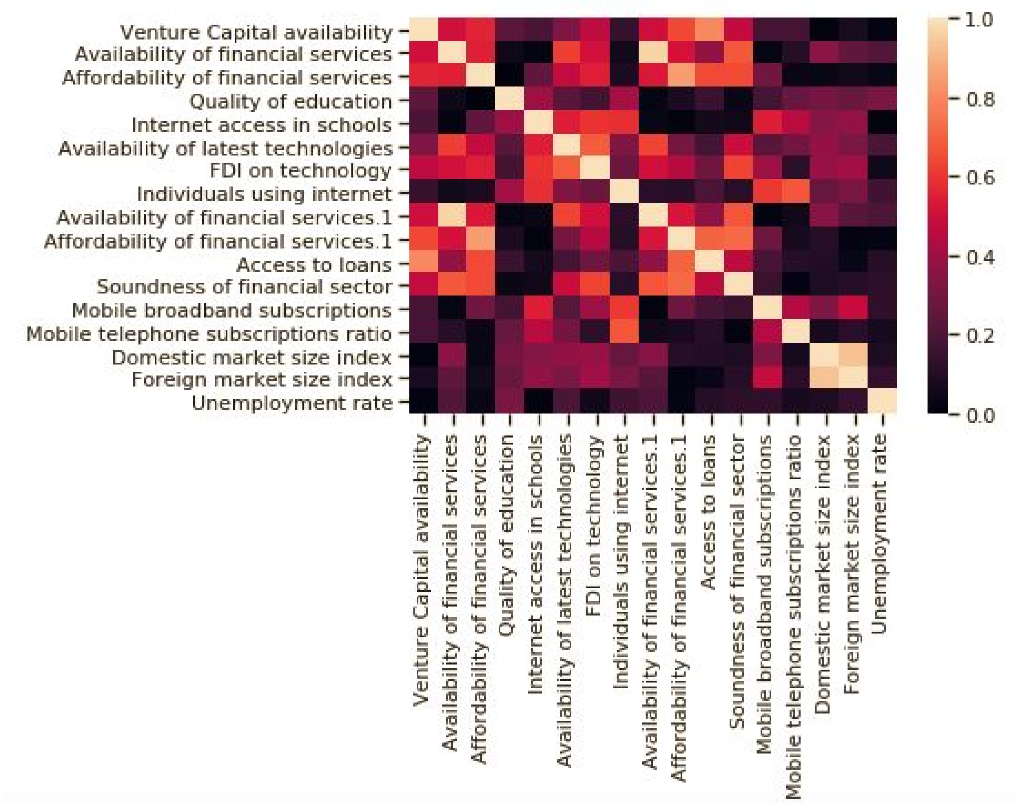




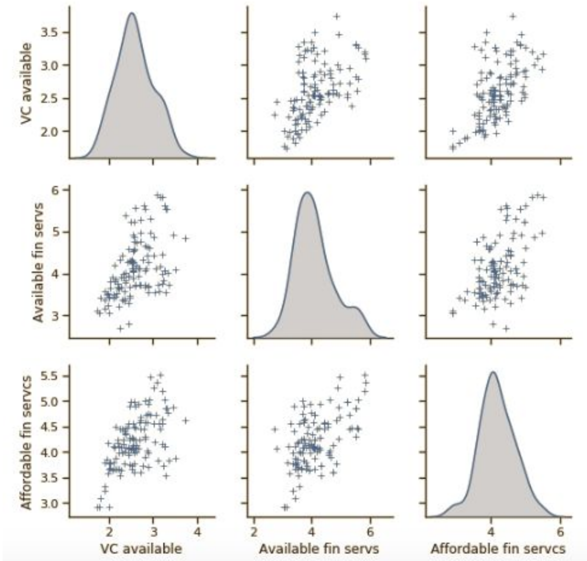
As seen below there is a growing tendency of new Fintech companies which peak in recent years for Latin American countries.

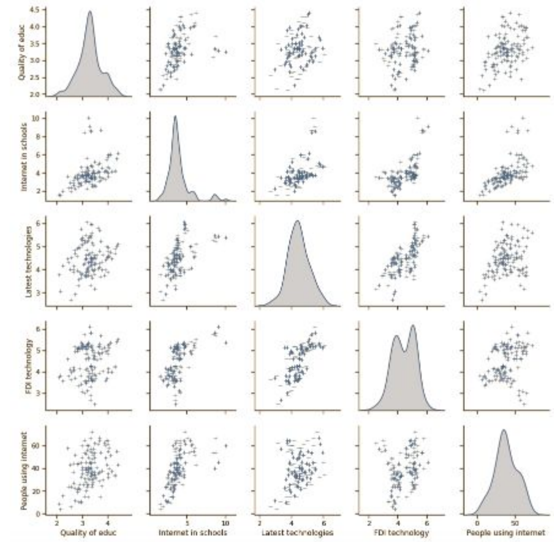


The correlation map shows interesting relations that will not be explored in depth.

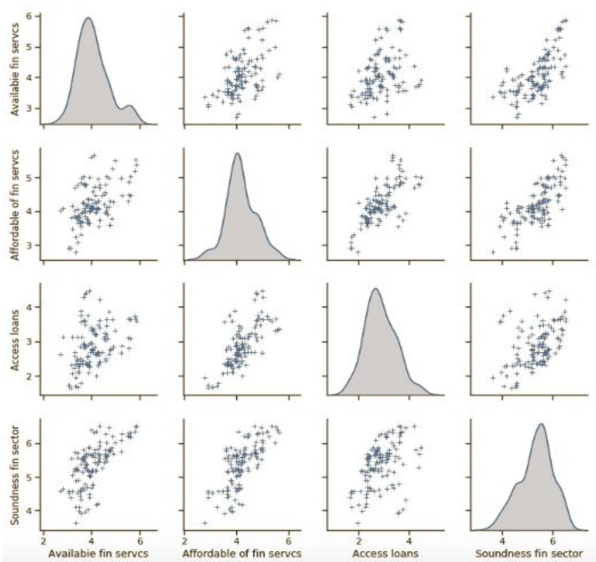


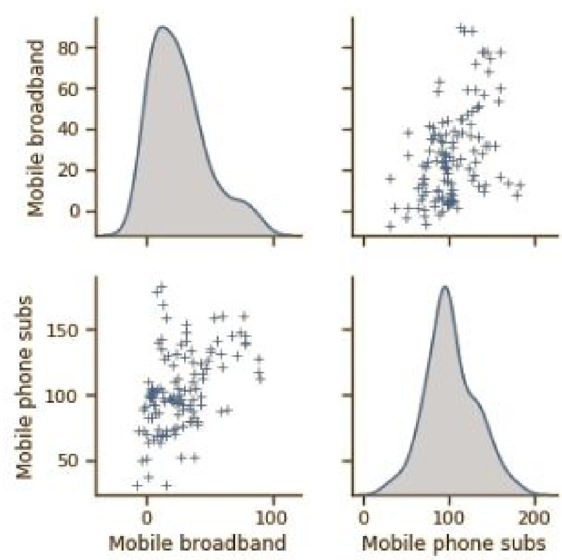
The graph on the left shows contain distribution and pairplots for the following capital development related variables in the following order: Venture capital availability, Availability of financial services, Affordability of financial services. The second one contains the same information for technology available variables: Quality of education, Internet access in schools, Availability of latest technologies, FDI on technology, and Individuals using internet.



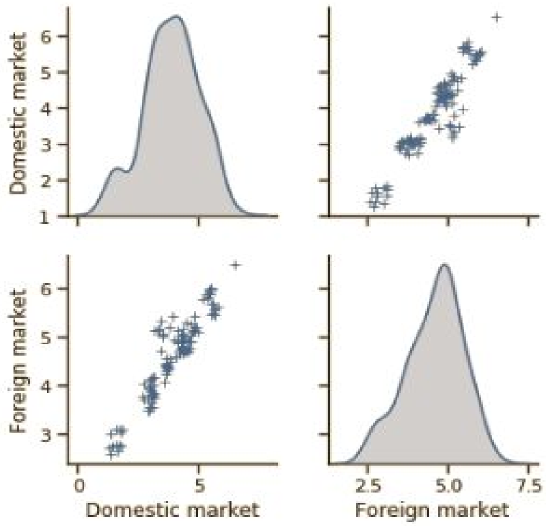


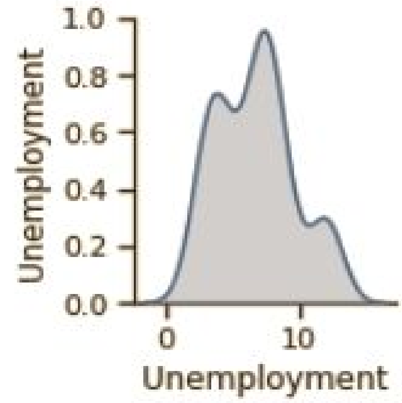
Below is the normality and pairplot relation among financial soundness variables (5) and mobile phone use variables (2).



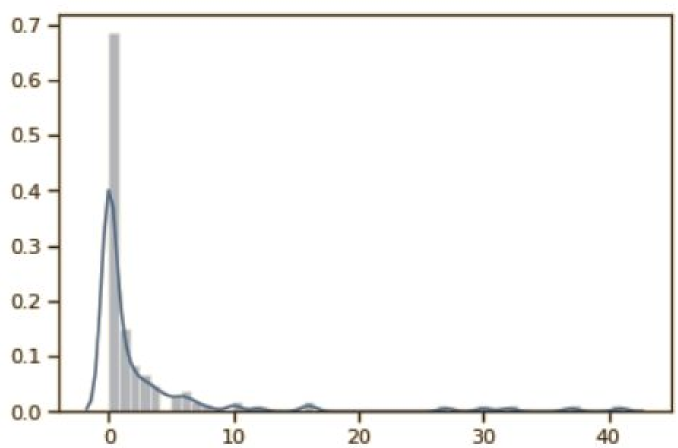


The next visuals is the normality and pairplot relation among financial market size variables (3) and the unemployment variable, which was extracted from the World Bank data.





The number of Fintechs per country for each year variable behaves similar to the international case studied in part I. It demonstrates that the data does have a normal distribution, and this time the mayor outlier is Brazil which consistently has a notably above average number of new Fintech companies each year.



Weighted variables

Weighted variables follow the same structure as in part I.

Weighted Target variable

For this part, not only the number of Fintechs will be contemplated as a target variable, but also the average Crunchbase ranking in that year will be considered. For this it there must be introduced two auxiliary variables that enable both concepts to have the same range of values.

Fintechs_decile:
Takes values from a range (0,1). It is an ordinary variable that takes the highest number of Fintech companies in a row as a 1 and the lowest number as a 0 and the remaining values take values from 0 to 1 in accordance to their ranking.

CB_decile:
Takes values from a range (0,1). It is an ordinary variable that takes 1 for the best Crunchbase ranking average of Fintech companies in a row, while 0 accounts for the lowest ranking and the remaining values take values from 0 to 1 in accordance to their ranking.

Master_Target:
This variable takes the weighted values of “Fintechs_decile” and “CB_decile”. In this case “Fintechs_decile” takes 80% of the whole variable and “CB_decile” takes the remaining 20% as follows:

Master_Target = (Fintechs_decile x 4) + CB_decile

Note direct relation between “Countries_fintech” and “Fintech_decile” variable and the indirect relation among “CB_ranking” and “CB_decile” (a high ranking means low quality and vice versa), whereas “Master Target” is the result of 4x “Fintechs_decile” plus “CB_decile”.

Countries_fintechs	CB_Ranking	Fintechs_decile	CB_decile	Master_Target	Country	Latam	Year
4	229	0.797	0.642	2.046696	Argentina	1	2018
32	300	0.976	0.593	2.315072	Brazil	1	2018
1	424	0.537	0.561	1.205028	Chile	1	2018
7	175	0.894	0.715	2.556840	Colombia	1	2018
6	176	0.862	0.699	2.410152	Mexico	1	2018

Preprocessing

Data preprocessing follows the exact same process as in part one where:

- Data standardization was performed for master variables with preprocessing module from sklearn Python package.
- Normality in the sample was assured, the *Normalizer* command from the preprocessing module of Python's sklearn.

The remaining dataset containing only master variables which are centered, scaled and normalized has the following first rows:

	Master_Target	Master_one	Master_two	Master_three	Master_four	Master_five	Master_six
count	130.000000	130.000000	130.000000	130.000000	130.000000	130.000000	130.000000
mean	-0.024455	-0.012779	-0.015403	-0.006703	-0.013165	-0.006263	-0.014799
std	0.443610	0.355034	0.356775	0.365946	0.356899	0.363020	0.403808
min	-0.833976	-0.707899	-0.708827	-0.772367	-0.659865	-0.727285	-0.763423
25%	-0.381084	-0.263427	-0.277221	-0.261839	-0.269088	-0.242225	-0.370242
50%	-0.242583	-0.032275	-0.050160	0.050909	-0.117549	0.014519	0.025031
75%	0.414759	0.271353	0.236364	0.262070	0.247888	0.265398	0.243196
max	0.890645	0.699971	0.755580	0.581655	0.770594	0.645470	0.931286

While the master target variable has a normal distribution.

The next table shows the correlation between master variables for the Latin American sample.

	Master_Target	Master_one	Master_two	Master_three	Master_four	Master_five	Master_six
Master_Target	1.000000	0.215843	0.431304	0.359351	0.383174	0.509423	0.106377
Master_one	0.215843	1.000000	0.386681	0.662622	-0.048118	0.124221	-0.014651
Master_two	0.431304	0.386681	1.000000	0.440742	0.467604	0.313826	0.153095
Master_three	0.359351	0.662622	0.440742	1.000000	0.067570	0.116716	0.051079
Master_four	0.383174	-0.048118	0.467604	0.067570	1.000000	0.051741	0.088565
Master_five	0.509423	0.124221	0.313826	0.116716	0.051741	1.000000	-0.006851
Master_six	0.106377	-0.014651	0.153095	0.051079	0.088565	-0.006851	1.000000

The hypothesis interpretation is as follows:

- **H0 or null hypothesis:** the two variables are independent.
- **H1 or alternative hypothesis:** there exists a linear dependency between the variables.

* Note that the following hypothesis are the same as in part one. However the geographic region studied is different. This is done so, to identify any unique characteristics of the socioeconomic factors that enable Fintech company emergence in Latin America in comparison to the international realm.

- **H1:** How capitalized an economy is, positively affects the emergence of new Fintech companies in the sample studied.
- **H2:** Available technology in countries foster the emergence of new Fintech companies in the sample studied.
- **H3:** Sound financial systems enable Fintech company emergence in the sample studied.
- **H4:** The mobile subscription total number is a determinant for the emergence of new Fintech companies in the sample studied.
- **H5:** A large market size of an economy relates positively to the emergence of new Fintech companies in the sample studied.
- **H6:** Higher levels of unemployment fosters emergence of new Fintech companies in the sample studied.

Method and assumptions are the same as in part I.[9]

Results

H1: How **capitalized an economy** is, positively affects the emergence of new Fintech companies.
p value: `0.013649375512200086`
ACCEPT ALTERNATIVE HYPOTHESIS

H2: Available technology in countries foster the emergence of new Fintech companies.
p value: `3.0082661727792637e-07`
ACCEPT ALTERNATIVE HYPOTHESIS

H3: Sound financial systems enable Fintech company emergence.
p value: `2.6833653524601116e-05`
ACCEPT ALTERNATIVE HYPOTHESIS

H4: Mobile subscriptions is a determinant for the emergence of new Fintech companies.
p value: `6.807286556366098e-06`
ACCEPT ALTERNATIVE HYPOTHESIS

[9] Method: one tailed Peason correlation test. Assumptions: normality of variables, independence of variables, and the same variance.

H5: A large **market size** relates positively to the emergence of new Fintech companies.
p value: `6.032682940493941e-10`
ACCEPT ALTERNATIVE HYPOTHESIS

H6: Higher levels of **unemployment** fosters emergence of new Fintech companies.
p value: `0.22835972790643308`
REJECT ALTERNATIVE HYPOTHESIS

Accepted alternative hypothesis ordered by level of significance:

Market size (Hypothesis 5) > technology (Hypothesis 2) > mobile phones (Hypothesis 4) > financial soundness (Hypothesis 3) > capital (Hypothesis 1)

Observations

- The same alternative hypothesis are accepted in the International case (part I) and the Latin American case (part II).
- Level of significance vary from both regions.
- Order of significance between regions differ.

Hypothesis 7 Determinants for Fintech expansion are increasingly favorable in recent years.

Methodology

All master variables that prove to have significance in order to explain linear dependencies with Fintech company emergence in the Latin American sample will be considered.[10]

A variable called score is built, where each of the variables have the same weight regardless of different levels of significance. Each of the five variables account for 20% of the weighted value of the score as follows[11]:

Score = (master1*0.2)+(master2*0.2)+(master3*0.2)+(master3*0.2)+(master4*0.2)+(master5*0.2).

Each datapoint is the result of the division of each value by the maximum value of that variable of column multiplication by two (in order to have 0 as the minimum value addressable and 10 the maximum value possible) as follows:

---

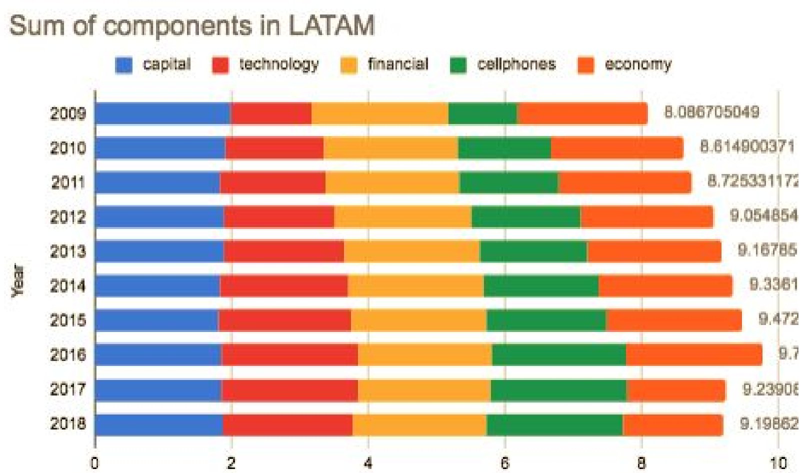
[10] master1 (capital), master2 (technology) , master3 (financial soundness), master4 (mobile phones), master5 (market size).
[11] All values from master variables already centered, scaled and normalized as in previous hypotheses in part II.

Datapoint = (x/(max value in column)) x 2

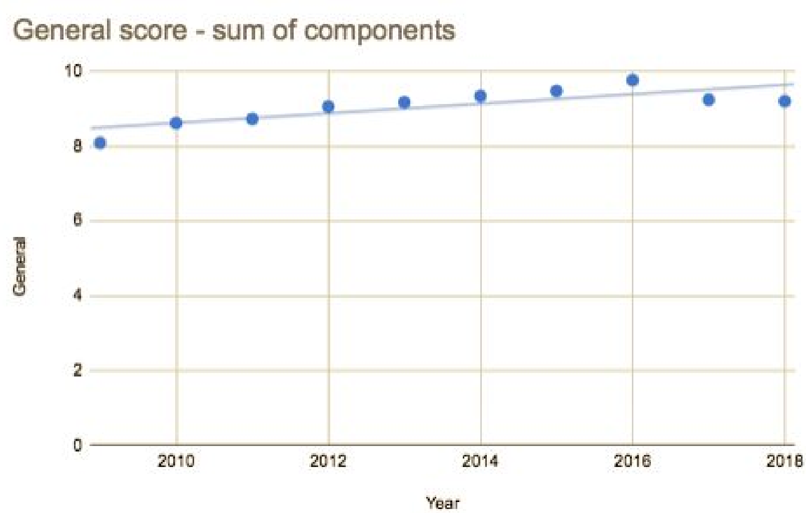
The resulting table is and graphs are as follows, where the column “General” is the horizontal sum of each datapoint:

Year	capital	technology	financial	cellphones	economy	General
2009	1.997309014	1.180095543	1.988182027	1.024323552	1.896794914	8.086705049
2010	1.913399907	1.430607921	1.968264065	1.36356059	1.939067889	8.614900371
2011	1.827337523	1.537778032	1.965718548	1.462135484	1.932361585	8.725331172
2012	1.889808149	1.619409856	1.999935444	1.607199029	1.938501804	9.054854282
2013	1.890781782	1.752921982	1.997658587	1.567353172	1.959135803	9.167851326
2014	1.837148197	1.870730597	1.980537775	1.679019377	1.968711651	9.336147597
2015	1.815008255	1.936164279	1.980537775	1.767919581	1.973084706	9.472714596
2016	1.861121136	1.993197924	1.949052446	1.960397757	1.999962296	9.763731558
2017	1.859338193	1.999448648	1.940640009	1.985947506	1.45370818	9.239082536
2018	1.875302663	1.901469225	1.948991893	1.999755614	1.473103169	9.198622564

The sum of each component is represented in the following graph.



The tendency shows a clear improvement in the general component score.



To verify statistically if there is statistical evidence of improvement in the score of Fintech socioeconomic enablers, there will be compared the first 5 years general score (2009-2013) against the last five years considered in the sample (2014-2018).

This time **Student's t-test** will be performed to test whether the means the means two independent samples are significantly different. To perform this test, the same assumptions of part I and II have to be covered, and are fulfilled.

The hypothesis interpretation is as follows:

- **H0 or null hypothesis**: the means of the samples are equal.
- **H1 or alternative hypothesis**: the means of the samples are unequal.

Result

Hypothesis 7: Determinants for Fintech expansion are increasingly favorable in recent years.
p value**:** `2.772827354775973e-05`
ACCEPT ALTERNATIVE HYPOTHESIS

Part III.
Fintech B2B-B2C market trend analysis

Dataset
"Mastersheet_part3.csv": shape 15 x 77

	Date	Quarter	B2B	B2C	Mean CB rank	CB rank percentile	Mean CB rank_B2B	Mean CB rank_B2C	Sum of both	First Half	Second half	Difference
0	2000-03-31	2000 Q1	3	6	390,421	0.333	158	239	9	1	0	-3
1	2000-06-30	2000 Q2	1	1	567675	0.250	567	567	2	1	0	0
2	2000-09-30	2000 Q3	0	0	316495.9863	0.833	312	289	0	1	0	0
3	2000-12-31	2000 Q4	0	0	316495.9863	0.000	312	289	0	1	0	0
4	2001-03-31	2001 Q1	1	2	552,223	0.286	440	422	3	1	0	-1

Source
Crunchbase

Variables
Date
Quarter
B2C
B2C
Mean CB rank
Mean CB rank_B2B
Mean CB rank_B2C
* other columns displayed in the table are operations from the variables listed above.

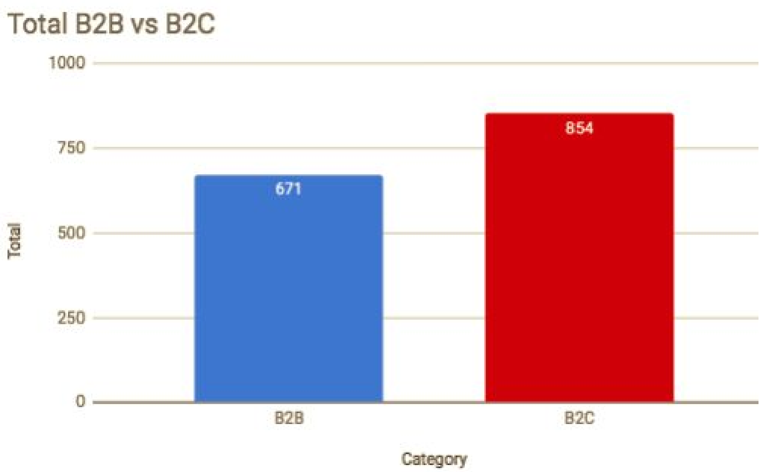
Time period
2000 - 2018

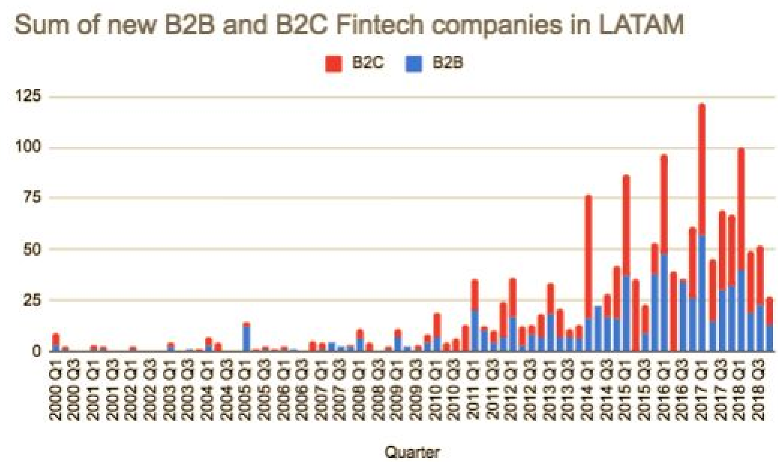
* Countries evaluated are the same as in part II (all South American countries and Mexico).

Completeness
Information retrieved was complete

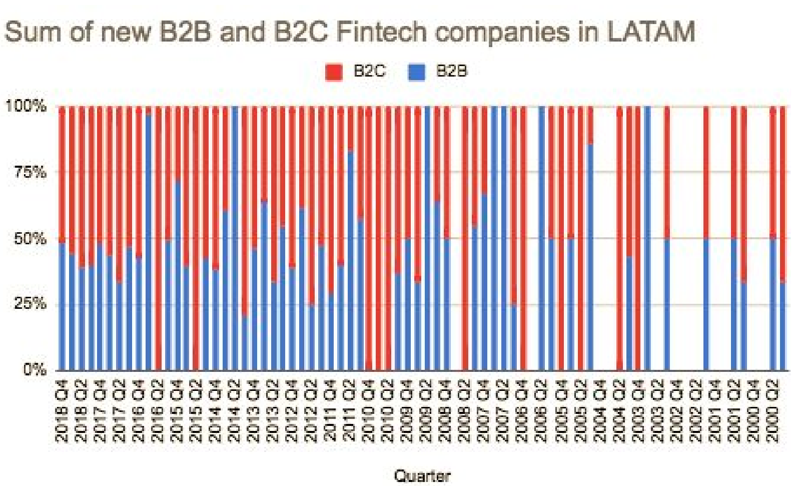
Exploratory Analysis part III

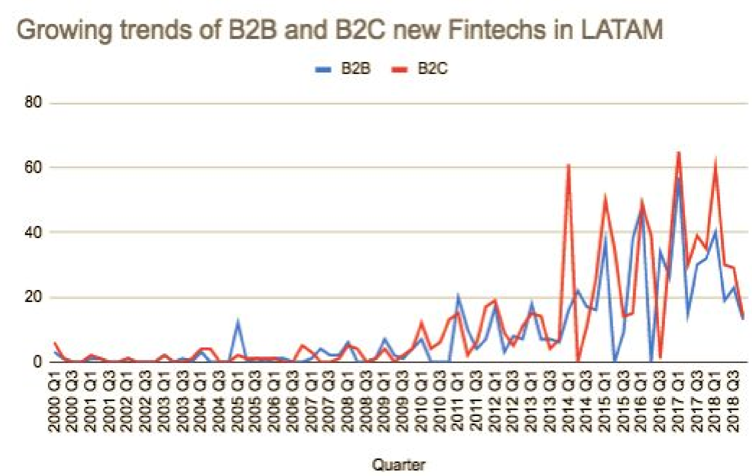
This section will explore valuable market insights and identify any trends regarding B2B and B2C Fintech companies in Latin America. The left graph demonstrates that there are more B2C new established companies in the period studied (2000-2018) in Latin America and Mexico. The right sided visual illustrates a growing tendency of Fintech founded companies reaching more than 100 new companies in the first quarter of 2017.



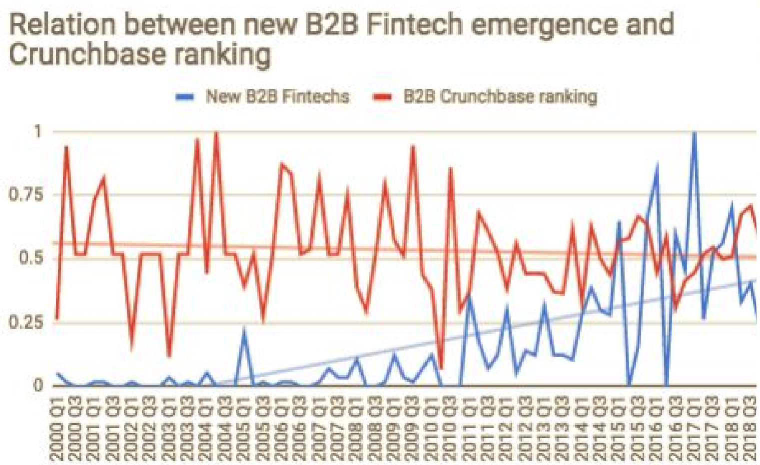


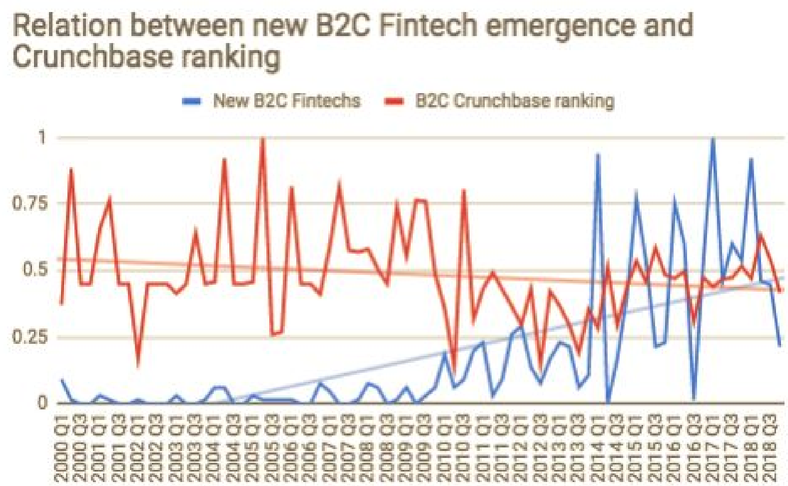
The first graph from left to right illustrates the share of B2B and B2C new companies in Latin America. In recent years the percentage has become similar for both categories, while in previous quarters the relation among both seem unstable because the sample was not numerous. The second graph illustrates a linear tendency between the number of B2B new companies and B2B new companies in the region studied.



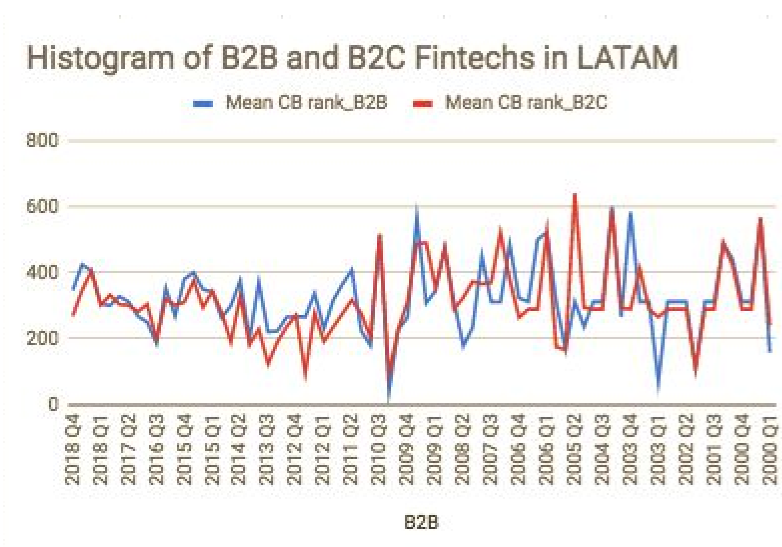


Left and right graphs show the relation between Fintech new companies and Crunchbase ranking of B2B and B2C respectively. For both cases, there is a notable and sustained increase in the number Fintechs. However, the quality of these companies, which is covered in the Crunchbase ranking, demonstrates that on average these companies don't have a better ranking.





The next visual demonstrates an arbitrary relation between B2B and B2C Fintech company emergence and time. Nevertheless, there seems to be a linear dependence of B2B and B2C ranking, this can be explained partially due to the fact that some companies are considered to be both B2B and B2C and thus have the same ranking.



Weighted variable

Two variables called “Master_B2B” and “Master_B2C” will be introduced. This variable will only be used to prove hypothesis 7 and it contemplates two concepts:

- Percentile of New Fintechs in a given time period. (“Decile_number_B2B” and “Decile_number_B2C” respectively)
- Percentile of Crunchbase mean ranking of the new Fintechs in question. (“Decile_rank_B2B” and Decile_rank_B2C “respectively”)

Master_B2B = (Decile_number_B2B x 4)+ Decile_rank_B2B

Master_B2C = (Decile_number_B2B x 4)+ Decile_rank_B2B

Hypotheses

Below are relevant hypothesis in terms of Fintech market B2B-B2C trends are the following that reduce ambiguity to illustrate how the sector is evolving in terms of quantity of new companies, composition, and quality.

- **H1:** Fintech new companies are increasing in Latin America.
- **H2:** The B2B number of new companies is increasing in the sample explored.
- **H3:** The B2C number of new companies is increasing in the sample explored.
- **H4:** The Crunchbase ranking of companies in the sample reduced overall.
- **H5:** There are significantly more B2C companies than B2B in the sample studied.
- **H6:** B2C Fintech companies has a better score than B2B in general.
- **H7:** The Score Gap between both categories has reduced in recent years.

The hypothesis interpretation is as follows:

- **H0 or null hypothesis**: the means of the samples are equal.
- **H1 or alternative hypothesis**: the means of the samples are unequal.

Methodology

Each Hypothesis contains two variables. The results required for the last 5 years of the sample studied and the results for the first 13 years explored.

For example, the first hypothesis compares the number of new companies in the last 5 years as opposed to the first 13 years of the sample, and the table has the following structure, where column "Last_5_B2C" records the last 5 years in quarter periods, while "First_B2C" records new Fintech companies in the time period 2000-2014:

	Last_5_B2C	First_B2C
0	11.0	12
1	23.0	8
2	24.0	6
3	56.0	21
4	40.0	13
5	37.0	5
6	25.0	6
7	69.0	23
8	36.0	7
9	20.0	3

This structure is followed for all the other 7 hypotheses of this section.

Data imbalance

If explored carefully the left sided column from the above table, has to have less data points since it only accounts information from a period of 5 years since the other column accounts for a total of 13 years. To fill this gap the remaining values were filled by taking the mean of the variable and summing and resting the standard deviation consecutively as follows.

- *Missing value a* = $\overline{X} + \sigma$
- *Missing value b* = $\overline{X} - \sigma$

E.g: *hypothesis x data* → $\overline{X}$ = 10, σ = 1

- *Missing values of hypothesis x* = 9, 11, 9, 11, 9, 11 …..

This way the sample retains its mean and standard deviation which will be used in the following hypothesis testing.

Testing normality

Normality test was performed with from scipy Python's package, taking relaxed p.value of normality test of 0.075. Below are the results of each variable taken in the hypotheses:

Interpretation

- **H0 or null hypothesis:** the variable does not present normality.
- **H1 or alternative hypothesis:** the variable presents normality.

Normality Testing Results of Variables

Last_5_both (H1): p = 3.12409e-05 → PRESENTS NORMALITY
First_both (H1): p = 3.75384e-09 → PRESENTS NORMALITY

Last_5_B2B (H2): p = 0.000152764 → PRESENTS NORMALITY
First_B2B (H2): p = 1.48977e-07 → PRESENTS NORMALITY

Last_5_B2C (H3): p = 6.06552e-05 → PRESENTS NORMALITY
First_B2C (H3): p = 1.77403e-08 → PRESENTS NORMALITY

Difference_last_5 (H4): p = 0.059588 → PRESENTS NORMALITY
Difference_first (H4): p = 1.84661e-10 → PRESENTS NORMALITY

Ranking_B2B_1 (H5): p = 2.91455e-75 → PRESENTS NORMALITY
Ranking_B2C_1 (H5): p = 1.7153e-251 → PRESENTS NORMALITY

Difference_1 (H6): p = 0.0119933 → PRESENTS NORMALITY
Difference (H6): p = 0 → PRESENTS NORMALITY

Diff_score_last5 (H7): p = 0.00209086 → PRESENTS NORMALITY
Diff_score_first (H7): p = 0.305928 → **DOES NOT PRESENT NORMALITY**

Normalization for not normalized variables

To assess lack of normality in the variable xxxx from H7, the *Normalizer* command from the preprocessing module of sklearn was used. In this case both "Diff_score_last5" and "Diff_score_first" from H7 were normalized since the command request that the whole dataset needs to be normalized. After this, the normality test showed successfully that both variables present normality as follows:

Diff_score_last5 (H7): p = 0.0123049 → PRESENTS NORMALITY
Diff_score_first (H7): p = 0.0212875 → PRESENTS NORMALITY

Since the variables that are contained in the hypotheses fit the requirement to be statistically tested, **Student's t-test** will be performed to test whether the means the means two independent samples are significantly different. The considered p value threshold is 0.05.

Results

H1: Fintech new companies are increasing in Latin America.
p value**:** `4.64374695733805e-26`
ACCEPT ALTERNATIVE HYPOTHESIS

H2: The B2B number of new companies is increasing in the sample explored.
p value**:** `1.6190689477759134e-27`
ACCEPT ALTERNATIVE HYPOTHESIS

H3: The B2C number of new companies is increasing in the sample explored.
p value**:** `4.066365380882651e-24`
ACCEPT ALTERNATIVE HYPOTHESIS

H4: The Crunchbase ranking of companies in the sample reduced overall.
p value**:** `0.08637712340702586`
REJECT ALTERNATIVE HYPOTHESIS

H5: There are significantly more B2C companies than B2B in the sample studied.
p value**:** `0.032567490372818024`
ACCEPT ALTERNATIVE HYPOTHESIS

H6: B2C Fintech companies has a better score than B2B in general.
p value**:** `2.1597589114147994e-06`
ACCEPT ALTERNATIVE HYPOTHESIS

H7: The Score Gap between both categories has reduced in recent years.
p value**:** `0.19781764713383684`
REJECT ALTERNATIVE HYPOTHESIS

Time Series Analysis for new Fintech companies in Latin America

The objective of this section is to execute a forecast of the development of the Fintech market in Latin American. The following aspect of the Fintech market for the sample studied will the predicted:

- How will the total number of new Fintech companies evolve in the next 3 years.

For this section Univariate Time Series Forecasting will be applied. The ARIMA algorithm stands for 'AutoRegressive Integrated Moving Average', will be used[12].

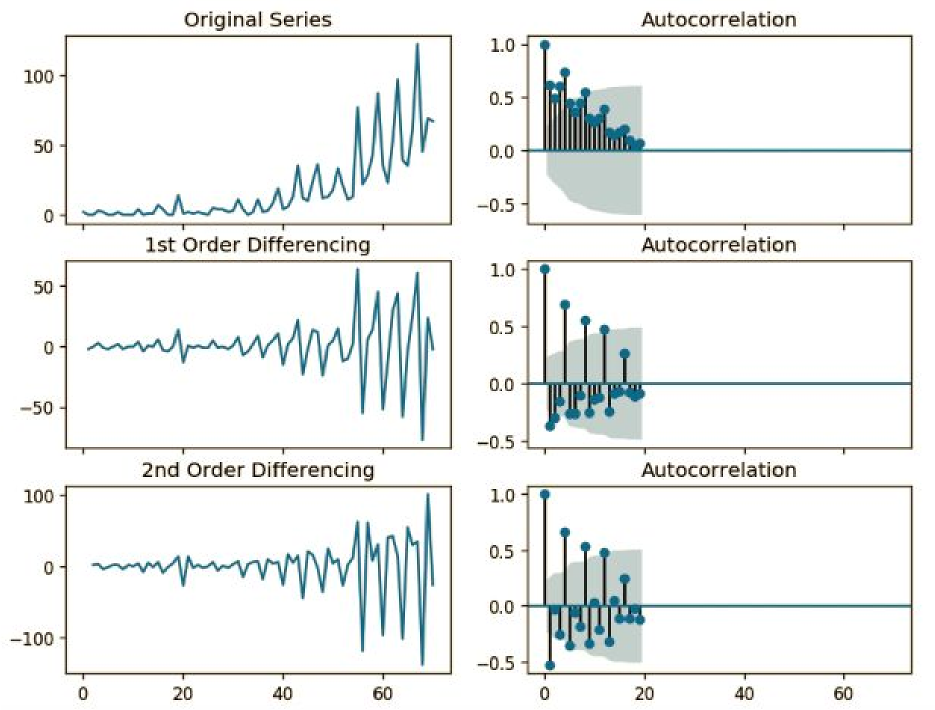
An ARIMA model is has 3 main parameters: p represents the order of the AutoRegressive term, q refers to the order of the Moving Average parameter, while d is the number of differencing to convert the time series into stationary.

Testing is series is stationary.

```
ADF Statistic: 4.311433
p-value: 1.000000
```

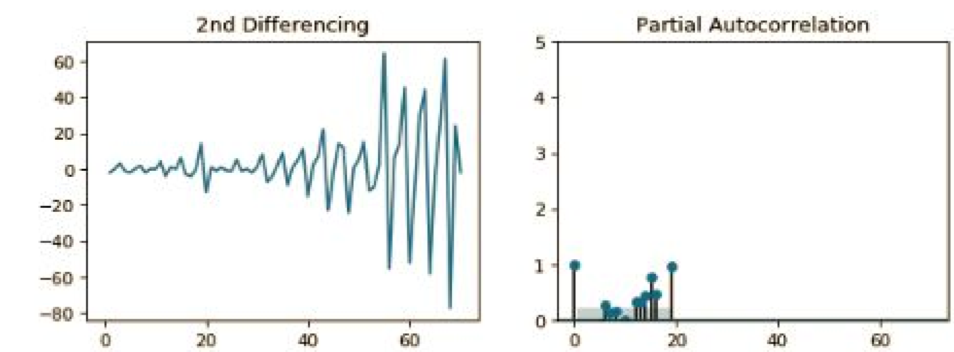
Since the p-value of the ADF test does not fulfill the significance level (0.05) then the time series is not considered stationary. In this case, there must be found the order of differencing.

[12] This algorithm is based on the idea that the past values of the time series can solely be used to predict the future values.



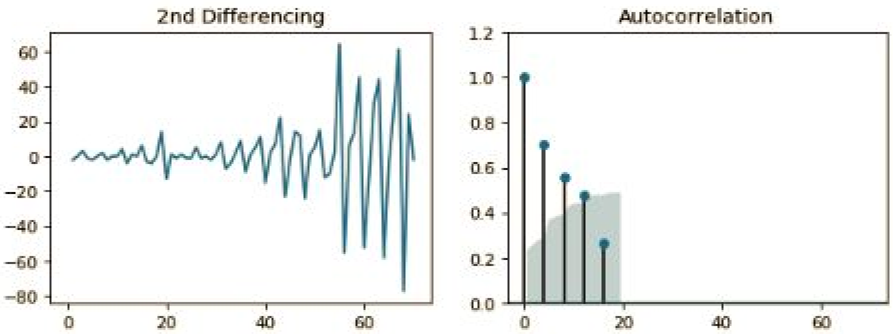
The time series has stationarity with both orders of differencing. This time the second order differencing (d) will be selected.

The following graph shows the *second order* differencing order series and the partial autocorrelation.



In this case the P term selected will be *1* for the AR (autoregressive term).

Next is illustrated the second differencing and autocorrelation of the second differencing.



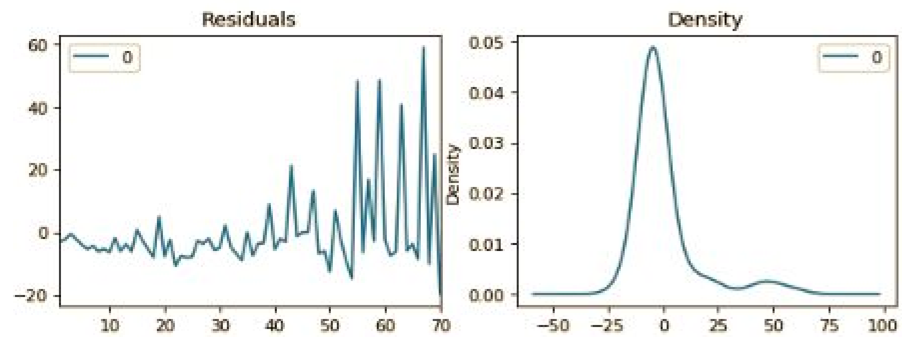
The q term selected for this model is 1.

The Arima selected model takes d value as 2,p as 1 and q as 1. That is portrayed in the model architecture below.

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                D.value   No. Observations:                   70
Model:                 ARIMA(2, 1, 1)   Log Likelihood                -285.184
Method:                       css-mle   S.D. of innovations             14.040
Date:                Sat, 07 Sep 2019   AIC                            580.369
Time:                        14:11:18   BIC                            591.611
Sample:                             1   HQIC                           584.834

=================================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
const             1.0086      0.337      2.989      0.004       0.347       1.670
ar.L1.D.value    -0.2803      0.111     -2.536      0.014      -0.497      -0.064
ar.L2.D.value    -0.5035      0.105     -4.805      0.000      -0.709      -0.298
ma.L1.D.value    -0.6550      0.075     -8.749      0.000      -0.802      -0.508
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -0.2783           -1.3815j            1.4093           -0.2816
AR.2           -0.2783           +1.3815j            1.4093            0.2816
MA.1            1.5266           +0.0000j            1.5266            0.0000
-----------------------------------------------------------------------------
```

To ensure there are no particular trends for residuals to ensure a normality by showing constant mean and variance plot will be executed as follows:



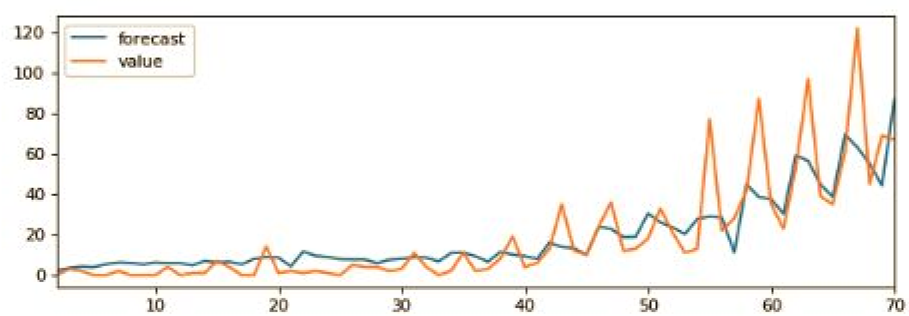
Residual errors seem behave with zero mean and uniform variance therefore, no transformation need to be done.

Model training and test split

The fist 85% data points will be used to dest the model, and the rest 15% will be used to test the model as follows:

train = df.value[:60]
test = df.value[60:70]

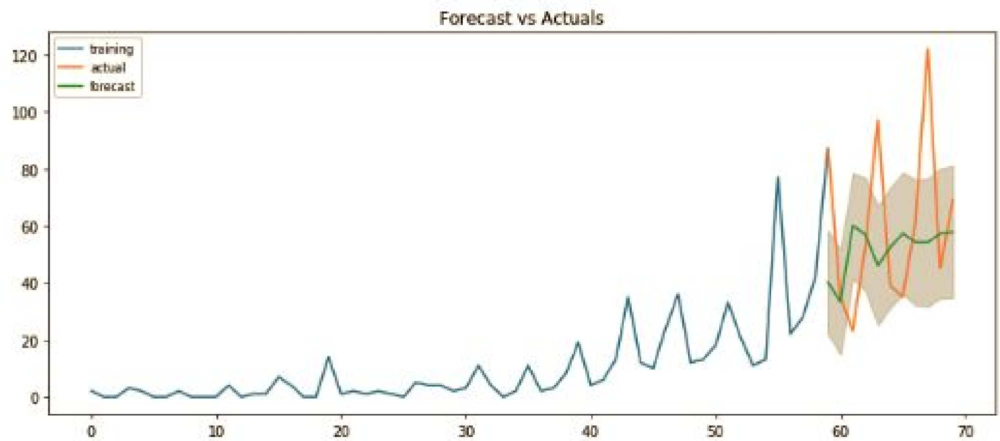
The following graph shows in the x axis uniform periods of time starting with the first quarter of the year 2000 and finishing with last quarter of 2018. With the selected model architecture the model forecasts behave in the following fashion:



As the graph above shows, the forecast follows a clear increasing tendency of new Fintech startups as the years pass. However sudden increases in recent years where notable local peaks appear do not necessarily are followed accurately by the model forecasts.

Validation model

This is a graphic representation of the validation model, which forecasts the results portrayed by the green line. As seen below, the sudden increase in the last years of Fintech was not accurately predicted by our model. That give evidence of a market boom of Fintech in Latin America in recent years.



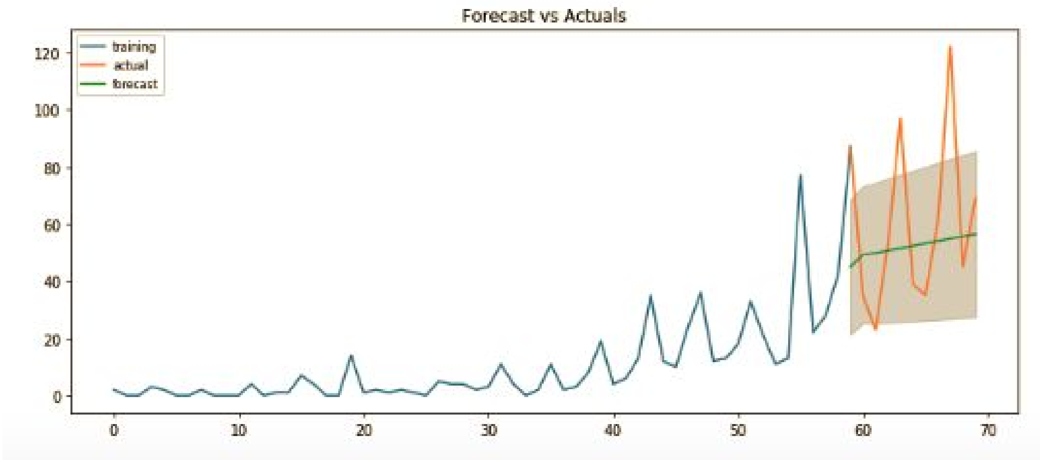
Modifying parameters

By testing in the programming framework and modifying parameter d=1 gave better results and reduced the AIC curve as seen in the next visual:

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                D.value   No. Observations:                   59
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -231.169
Method:                       css-mle   S.D. of innovations             12.074
Date:                Fri, 06 Sep 2019   AIC                            470.339
Time:                        22:32:26   BIC                            478.649
Sample:                             1   HQIC                           473.583

=================================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
const             0.8280      0.387      2.138      0.037       0.069       1.587
ar.L1.D.value    -0.0783      0.172     -0.456      0.650      -0.414       0.258
ma.L1.D.value    -0.7541      0.097     -7.739      0.000      -0.945      -0.563
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1          -12.7775           +0.0000j           12.7775            0.5000
MA.1            1.3260           +0.0000j            1.3260            0.0000
-----------------------------------------------------------------------------
```

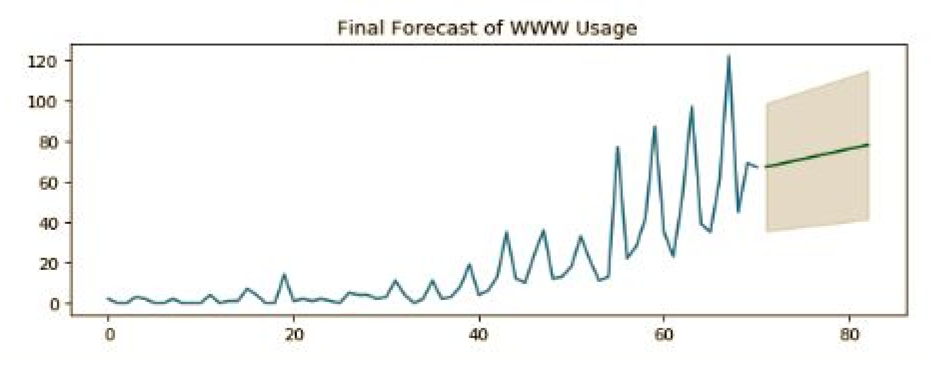
Validation model



This model performs better with first order differencing. The parameter selection of models is as follows: d = 1, p = 1, d = 1

Forecast

As presented in the beginning of this section, the aim of this segment is to predict how the total number of new Fintech companies evolve in the next 3 years. The model created visually presents the following trend, where the green line represents the predicted number of new Fintech companies for the next three years with a 95% interval of confidence covered by the gray area.



The array displayed is the following:

([67.13875436, 68.13077966, 69.12280496, 70.11483026, 71.10685556,
72.09888086, 73.09090617, 74.08293147, 75.07495677, 76.06698207,
77.05900737, 78.05103267])

This translates to the following time periods:

2019 Q1: 67.13875436
2019 Q2: 68.13077966
2019 Q3: 69.12280496
2019 Q4: 70.11483026
2020 Q1: 71.10685556
2020 Q2: 72.09888086
2020 Q3: 73.09090617
2020 Q4: 74.08293147
2021 Q1: 75.07495677
2021 Q2: 76.06698207
2021 Q3: 77.05900737
2021 Q4: 78.05103267

Mean of new Fintech companies forecasted in next three years: 72.59
Mean of new Fintech companies contained in the sample (2000-2018): 18.14

There is a **544.5% increase** in the mean of companies expected for the next three years 2019 - 2021 compared to the year period 2000-2018.

Discussion of results

The accuracy in terms of ‘RMSE’ is 30.290729784714063 which equated to 25.25% of the maximum value of the sample. The ‘RMSE’ of the model selected is 65% larger than the mean of the sample which is 18.14. This is evidence that the sector is impressively volatile and is experiencing a boom in recent years which represents high levels of complexity in order to model and forecast.

6. Conclusions and Future Work

6.1 Conclusions

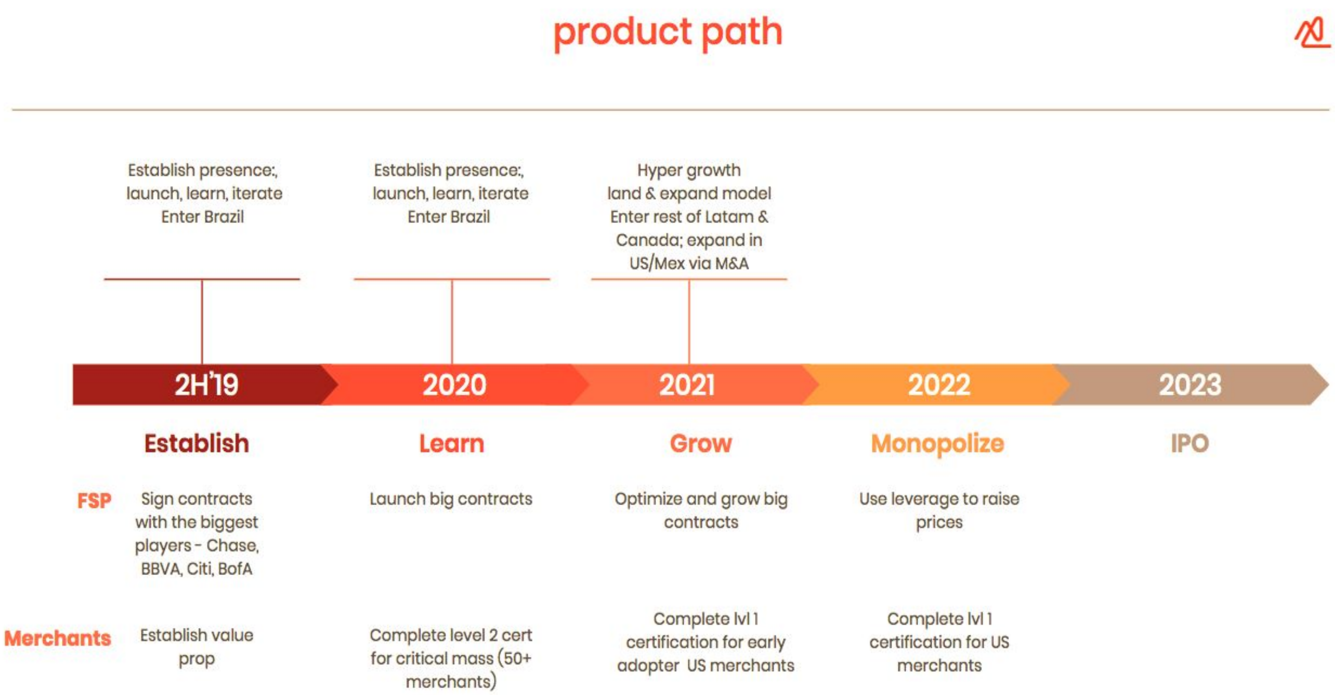
When tested statistically, the same alternative hypothesis were accepted in the International case (part I) and the Latin American case (part II). However, the level of significance varied for both regions, this may have many academic explanations that exceed the scope of this work. The components of Fintech emergence proved to be statistically significant displayed increasingly favorable results when compared to previous time periods, this translates to an increasingly favorable environment for Fintech company establishments in recent years.

When analyzing market trends for Fintech in Latin America, B2B and B2C companies, have consistently increased in number and the tendency shows a consistent increase of new companies in the industry. Time series analysis results forecasting give evidence that the sector is impressively volatile and is experiencing a major boom today.

6.2 Future work and application

In order to understand the implications of this work, it is important to clarify what the vision and route of the company is for the next five years. By the end of 2019 Arcus seeks to maintain contracts with banks such as BBVA, Chase, Citi and Bank of America, by 2020 it aims to establish its presence in the Fintech market and sign with more merchants and successively maintain its expansion to eventually become public in an IPO in 2023 as shown in the following visual.

Figure 4: Arcus 5 year route



extracted from Arcus company presentations

In this context, it is important that this work is inserted into the company's structure where it can have a high potential impact in the following five areas (1) market exposure, (2) first machine learning efforts, (3) raise funding, (4) showcase machine learning capabilities, and (5) open gate with the UK, which will be explained in the following paragraphs:

Press release: exposure for the company

One of the short term goals of the company, which are established as objective key results (OKRs) is to increase the company's exposure in the market. A possible application that have been thought about this work and that has been discussed with high-level personnel such as the Customer Relations Operator is publishing this study in a reduced version in specialized magazines such as The Fintech Weekly Magazine.

This would be an unprecedented document in Fintech. Although there have been written similar academic papers, this focuses its efforts on Fintech that operate B2B and seek to expand its market in Latin America. It is sought that through social networks specialized in businesses such as Linkedin the company officially publishes this work. In particular, this dissertation seeks to expose Arcus's bets on where the next market trends will be in the years to come. This work has the power to persuade that investors and other actors must bet on B2B Fintechs which have a presence in Latin America.

Baseline work: ground for future use cases and pattern recognition in the market

This work serves to establish bases for Machine Learning and Data Mining applications in the company. As such it is the first project of the company that involves Machine Learning Algorithms and practices. It is important to note that one of the strategic objectives of the company is to invest in Machine Learning to optimize its services (there are currently two projects in the area of customer success and business development) that can take this work as a foundation.

Raise funding: pitching for Series B

Since August, the company's CEO, Edrizio de la Cruz, will be centered exclusively in fundraising. The importance of this work is that it implies an empirical support of trends that he seeks to explain through his pitches, which are two: (1) the recent economic growth of the Fintech B2B market and (2) the Fintech boom in recent years in Latin America.

Subject to approval by the company management structure, fractions of this study will be shared in meetings with venture capital firms and different investors to give empirical evidence that the company's vision is supported by large databases and Machine Learning Analysis.

Showcase Machine Learning capabilities

Arcus recognizes the importance of staying at the forefront of technological capabilities to compete in the market. The importance of Machine Learning in this aspect resides in its ability to predict read market trends, automize business procedures and explore new product use cases. On the other hand, the company also considers it relevant to inform its potential investors that the company is already developing practices and applications related to Data Science.

Open gates with the UK

The United Kingdom is one of the most developed Fintech markets worldwide. It is not only interesting to think about providing services in the United Kingdom, but Arcus could build alliances and constructive relationships with British Fintechs, FSPs, merchants or investors. In addition, Arcus recognizes Goldsmiths University of London as a very prestigious and avant-garde university, and it will never discard to maintain contact with the Computing Department to exchange ideas or proposals for future projects.

In particular, the MS Data Science program has graduated professionals who are placed in the financial sector and in the business world that the company could keep in touch with. Finally, Arcus recognizes the talent in Goldsmiths University of London and would like to hold conversations about possible exchange internship programs for his students.

7. Bibliography[13]

1. “15 Statistical Hypothesis Tests in Python (Cheat Sheet).” *Machine Learning Mastery*, 8 Aug. 2019, machinelearningmastery.com/statistical-hypothesis-tests-in-python-cheat-sheet/.

2. Alois, JD. “3 Biggest Fintech Trends Shaping Latin America.” Crowdfund Insider, 9 June 2019, www.crowdfundinsider.com/2019/06/148254-3-biggest-fintech-trends-shaping-latin-america/.

3. “ARIMA Model - Complete Guide to Time Series Forecasting in Python: ML .” *Machine Learning Plus*, 28 Apr. 2019, www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/.

4. “The Berkus Method – Valuing the Early Stage Investment.” *BERKONOMICS*, berkonomics.com/?p=131.

5. Fontinelle, Amy. “What Exactly Is a Startup?” *Investopedia*, Investopedia, 14 July 2019, www.investopedia.com/ask/answers/12/what-is-a-startup.asp.

6. Haddad, et al. “The Emergence of the Global Fintech Market: Economic and Technological Determinants.” *SSRN*, 27 Aug. 2016, poseidon01.ssrn.com/delivery.php?ID=654125087110031107002006089026029076025033067063006028024025076018103083121105108006025011001016026124055019092071100105117091117023037077029103070025102118112079060075054120121069003000121016027095075086125110015026079029089096077064065093126123021&EXT=pdf.

7. “IPOs.” *Investopedia*, Investopedia, www.investopedia.com/ipos-4689824.

8. Jain, Nirvikar. “Fintech Valuation Methods for Money Transfer Disruptors.” *Toptal Finance Blog*, Toptal, 19 Mar. 2019, www.toptal.com/finance/valuation/how-to-value-a-company.

9. Jaksic, et al. “The Future of Banking: The Role of Information Technology.” *SSRN*, 16 Oct. 2015, poseidon01.ssrn.com/delivery.php?ID=616064002106075071121016115023112121037016025093044007014073016026002103074070123120028062030124045033010026076087080099124114046083078061083102071031066084124001000030016120077066066114106115070116101092122099094098121101110122125067075089096007125

[13] Style: MLA 8th Edition (assisted with EasyBib application from Google)

&EXT=pdf.

10. Mearian, Lucas. “What Is FinTech (and How Has It Evolved)?” *Computerworld*, Computerworld, 18 Sept. 2017, www.computerworld.com/article/3225515/what-is-fintech-and-how-has-it-evolved.html.

11. Nordqvist, Christian. “What Is a Unicorn Company? Definition and Examples.” *Market Business News*, Market Business News, 19 Feb. 2019, marketbusinessnews.com/financial-glossary/unicorn/.

12. “Perspectives.” *Greylock Perspectives*, news.greylock.com/perspectives/home.

13. Reiff, Nathan. “Series A, B, C Funding: How It Works.” *Investopedia*, Investopedia, 25 June 2019, www.investopedia.com/articles/personal-finance/102015/series-b-c-funding-what-it-all-means-and-how-it-works.asp.

14. Robehmed, Natalie. “What Is A Startup?” *Forbes*, Forbes Magazine, 15 May 2015, www.forbes.com/sites/natalierobehmed/2013/12/16/what-is-a-startup/#2215478f4044.

15. “Unicorns.” *Investopedia*, Investopedia, www.investopedia.com/unicorns-4689822.