NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Faculty of Computer Science
Bachelor's Program "HSE and University of London Double Degree Program in Data Science and Business Analytics"

Software Project Report

on the topic: Development of a Software Solution to Analyze Companies That Have Imposed Restrictions or Left the Russian Market in Response to Sanctions 2022

(interim, the first stage)

Fulfilled by the students of the groups БПАД-213 and БПАД-214:

Melissa Dibé, Oganesyan Tigran, Suchkov Ilya — БПАД-214

Son Ekaterina, Lyakhova Daria, Arkhipova Alexandra, Kruglikov Gleb — БПАД-213

__13.02.23__
Date

Checked by the Project Supervisor:

_____
Surname, First name, Patronymic (if any), Academic Title (if any)

_____
Job

_____
Place of work (Company of HSE Department)

Date _____ 2023 _____ _____
Grade (in 10-point scale) Signature

Moscow, 2023

Introduction

Relevance:

Many companies left Russia in 2022 due to the economic and political uncertainty in the country. The introduction of a range of restrictive legislation has made it increasingly difficult for foreign companies to do business in Russia. Additionally, sanctions imposed by the United States and other Western countries have had a detrimental impact on the Russian economy, making it even more challenging for businesses to to generate a profit.

Moreover, the pressures of customers and shareholders forced many international enterprises to withdraw their assets, liquidate outlets, and cease trade in Russia. Some businesses have altered their strategies and chosen to leave the country entirely after first adopting precautionary steps. Corporations also had reacted in plenty of other respects, such as by eradicating content and lowering the internet visibility of state-controlled news sources, which are owned by the Russian government. It is worth mentioning that several organisations that have started the process of leaving Russia have disclosed the negative financial impact on their companies.

Goals:

In this regard, we have decided to make a research that will be a comprehensive tool to track the companies that left Russia in 2022, their intentions and their actual behaviours towards the Russian Federation. The research paper would consist of data from various sources such as news reports, official documents, company websites, interviews with company representatives, etc. It would also use advanced analytics to identify patterns and trends in the data, such as Machine Learning to track the moods of those companies.

Moreover, the research would cover information on each company that left Russia in 2022. This information would include the reasons for leaving, the destination countries they chose, the effects of their departure on the Russian economy, their current relationships with Russia, and their future plans.

Two major groups would implement two different approaches to analyze crucial self-parsed database. Both of them would use Machine Learning, but one will use clustering methods and algorithms, while other Recurrent Neural Network (for more details visit the last chapter "Next Steps").

Basic terms and definitions

Database — a structured collection of data, generally stored and accessed electronically from a computer system. Databases store large amounts of information which can be used for various purposes such as analyzing trends, making predictions and storing records.

Hardware — physical components of a computer system, such as the monitor, mouse, keyboard, hard drive, and other parts. It can also refer to the computer itself.

Machine learning — a type of artificial intelligence (AI) that enables computers to learn from data without being explicitly programmed. It focuses on the development of computer programs that can access data and use it to learn for themselves. The process of machine learning is based on algorithms that iteratively learn from data and improve the accuracy of their predictions over time.

Cluster analysis — a type of unsupervised machine learning algorithm that groups data into clusters based on their similarity. It is used to explore and understand the structure of data by analyzing the relationships between different groups of data points. It can be used to find patterns in large datasets, identify outliers, and discover meaningful subgroups or clusters in the data.

Artificial Intelligence — is the ability of a computer or machine to think and learn. Examples of AI include voice recognition, problem-solving, learning, planning, and natural language processing.

Neural network — a type of machine learning algorithm inspired by the structure and functions of biological neural networks. Neural networks consist of layers of interconnected nodes that transmit signals from input data to an output. Each node is connected to multiple nodes in the next layer and receives signals from the previous layer.

Sanctions — form of punishment or restraint that is imposed by one country or international organization on another country as a result of the latter's violation of an agreement, law, or set of principles.

Models, algorithms and methods

During this period our team has been working on collecting the list of companies that have left Russia in 2022. To collect data on companies, we used web parsing techniques to scrape the relevant websites for information about these companies. In our case this included going through news web-pages and blogs. As our next step, we are going to scrap government websites for data on business registration records, as well as other sources such as social media for information about companies that have recently relocated outside of Russia. We are as well planning to use parsing to extract data from Twitter and LinkedIn to gain insights into why exactly companies may have decided to leave Russia.

Here is the list of the models, algorithms and methods that are to be used in the project:

Python programming language: The language that is widely used in scientific calculations, data scraping and analysis, machine/deep learning and offers a huge range of built-in tools.

Methods of web-parsing: Methods used to collect data from the internet by pulling out data from HTML/ XML files. Includes Python built-in libraries (such as requests) and additional open-source libraries (such as BeatifulSoup).

Methods of data analysis: Data analysis is done with the help of Python and functions from open-source libraries: Numpy and Pandas for data engineering, Matplotlib and Seaborn for data visualization.

Models and algorithms: For data clustering algorithms from scikit-learn Python library will be used (such as KMeans, DBSCAN, hierarchical clustering). To analyze tone of companies statements we will use Recurrent Neural Networks(RNN) models. They are implemented with use of tensorflow and pytorch (Python libraries for building Neural Networks and make tensor calculations) or specific libraries with pre-trained RNNs (such as IBM's "Tone Analyzer)"

Link to our main database: https://docs.google.com/spreadsheets/d/1PiUn5AZtbIO2Bo_1Y3cpcxf-YlYtakbw9jun4TIra8w/edit#gid=0

Contributions:

Dibé Melissa

Was working with the web-page called **InkoiN** — Russian blog about earnings, technology, self-development and finance.

Link to an article: https://inkocoin.ru/spisok-kompanij-ushedshih-iz-rossii-2022-iz-za-ukrainy-i-sankcij/

Link to the list of companies collected from the article: https://docs.google.com/spreadsheets/d/1Xome83CCQHdlq5GNHW1nT38cRMmsdh4Q4sb3ikVTnsk/edit#gid=0

Link to the parsing code: https://github.com/thisismemeli/InkoiN

In this specific article contains a list of corporations that have either completely departed the country or put certain constraints on the Russian market.

There were 354 firms mentioned, the vast majority of which were from the United States of America (chart 1) and the minority from Norway, Denmark, Taiwan and Canada. Furthermore, after studying the firms mentioned in this report, it was determined that the biggest part of those companies were from the Sports sector, the Clothes sector was in second place, and the Car market rounds out the top three (chart 2). Furthermore, the majority of companies have just ceased delivery to Russia. A lower percentage of the corporations have entirely stopped operating.

I as well used a few formulas and scripts to help me automate the process. For example, I used a script to search for specific words. This helped me save a lot of time when it came to filling out the database. And, in general, parsing was quite a useful skill and made the whole task much more efficient. To make the initial database in google sheets more accurate some manual work was done, to be more specific — I had to adjust some of the restrictions with accordance to the official statements of companies since some data in the article was a bit outdated.

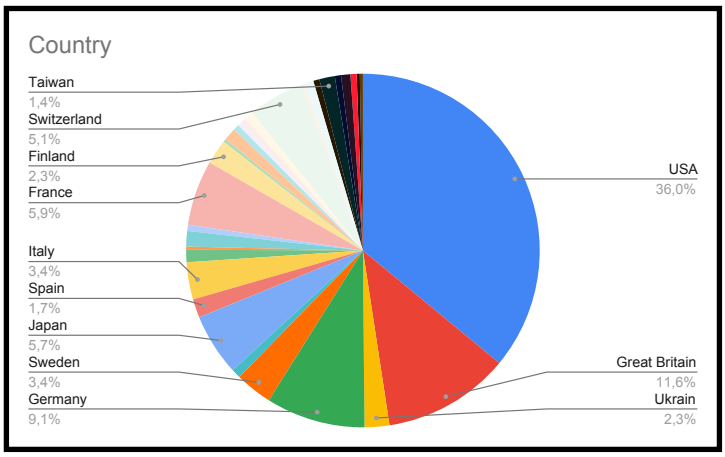


Chart 1 — Origin of companies.

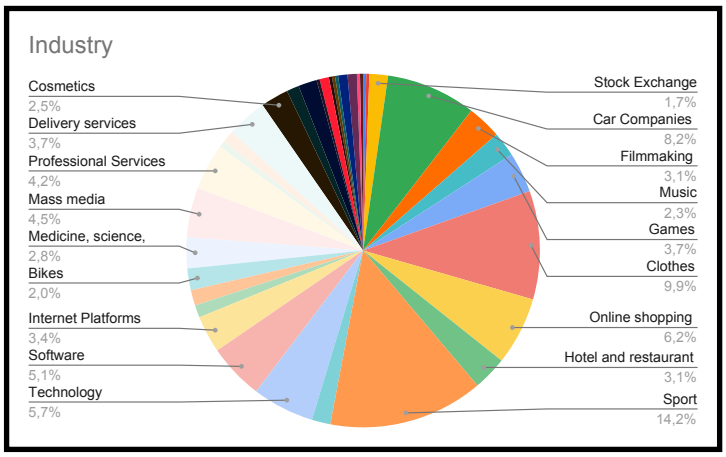


Chart 2 — Industry of companies.

Lyakhova Daria

Since 24 February many foreign companies have left Russia due to the military conflict between Ukraine and Russia. Many firms considered it inappropriate to continue operating on the Russian market due to the ethical part and their reputation or due to the pressure from the West. That is why many sites and articles gather information about firms, their parent countries and their field of activity. But all these sites do not have comprehensive information about it (some sites are not updatable, some have only partial information). The goal of our project is to collect relevant data from multiple resources, join it together and then analyze it on the basis of clusterization and the waves of withdrawal from the Russian market.

My part of the project was to collect all the data about companies that left Russia from the website called “Banki Today” (You can follow this link to view the site page). Banki Today portal is an independent financial online publication where anyone can find all the information they need on the most pressing issues: how to choose and apply for a loan, which deposit brings more income, which bank you can get a mortgage, and so on.

In order to gather information from the site page, I wrote the code that you can view by the link. To parse the site I used libraries such as BeautifulSoup, requests, and pandas:

1. Python software called Beautiful Soup is used to parse HTML and XML texts. For processed pages, it generates a parse tree that may be used to extract HTML data for web scraping.
2. For the Python programming language, there is an HTTP library called Requests. The project's objective is to make HTTP requests easier and more user-friendly.
3. Pandas is a data analysis and manipulation software package created for the Python programming language. It includes specific data structures and procedures for working with time series and mathematical tables.

Kruglikov Gleb

#Все**Запомним.рф**

Link: https://xn--80adjigxbghjs.xn--p1ai/

Web—Page:

This site was done by one of the enthusiastic user of Pikabu. With the news about companies leaving Russia and after multiple comments of Pikabu users he decided to make this little site-directory, where everyone can get info about all nowadays suspensions, bans and restrictions.

These days site have more than 700 companies, that imposed restrictions or left the Russian market, in the list. Every day it's updates by its own users (average +4 companies a day). Moreover it provides data about analogues for some companies.

Statistic:

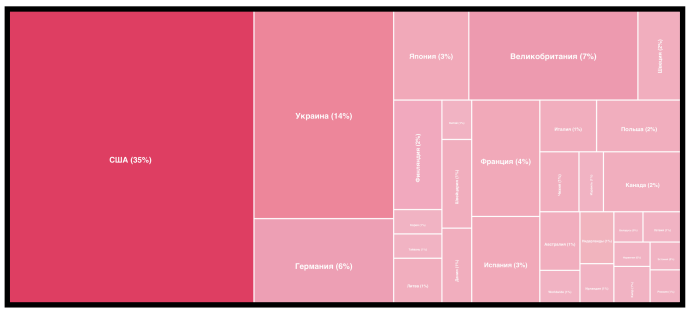


Chart 1 — Origin of companies.

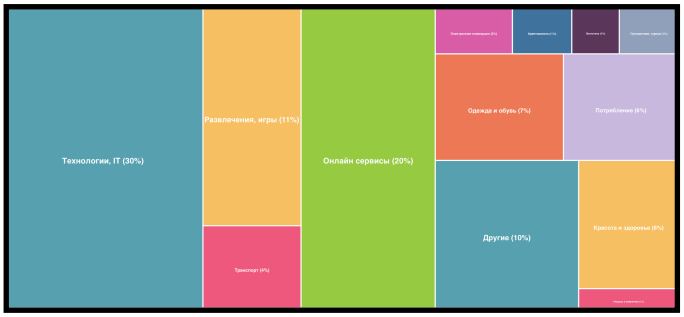


Chart 2 — Industry of companies.

As we can see top sector is IT with online services right after, moreover the top country is USA (correlation is obvious). I can suppose, that is because of the number of sanctions from this country also.

Progress of the work:

During taking out the information I faced some problems. First of all, this site do not provide official API or any file with all information. The solution wasn't delicate enough, but it worked. I used XHR (XMLHttpRequest) requests files. As I was looking for more companies, every time site was sending this type of file. After saving them into one big file and rewriting it as a python list with dictionaries inside I made it real to be parsed by python script and loading it into csv file. After all just a little corrections by hands and finally uploading to the joint base.

Code and text file with information about companies: https://github.com/Azas1K/Companies-leaved-russia

Used technologies:

1. **Python parser** I was looking forward parse the site through the official API and help of the Python library **BeautifulSoap.** But these approach wouldn't work as have told, there wasn't official IPA.
2. **XHR files** Therefore I had to do a little research about how I can reach the data without IPA. The answer was in XHR files, that site was sending while updating. Looking through them I found that it looks similar to the python dictionary, therefore after some manipulations it was made as dictionary, that I can already parse with ease by simple loop
3. **CSV** In the end I used CSV library to create a table with all the data.

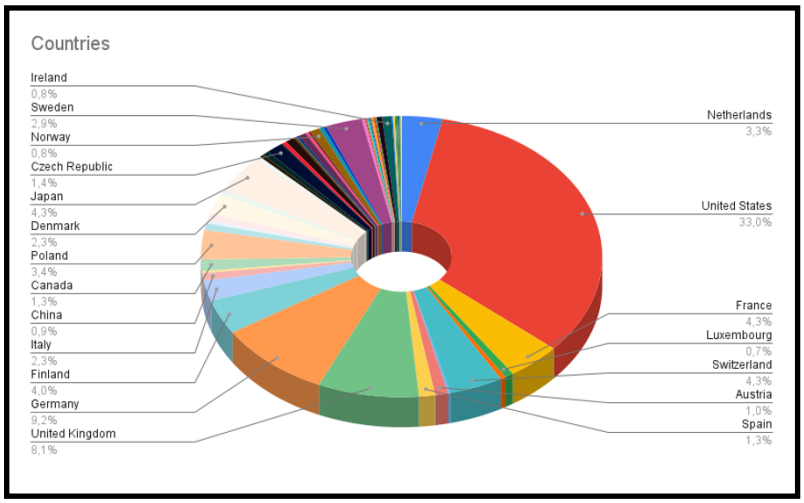
Oganesyan Tiran

Was working with the “Yale CELI List of Companies Leaving and Staying in Russia”.

Link to the webpage: https://www.yalerussianbusinessretreat.com/

Link to list with collected data: https://docs.google.com/spreadsheets/d/1Jm2_aL6JLUgiYrxHc4MRbYc2EjurKb5AW2J9EqfGdVs/edit?usp=sharing

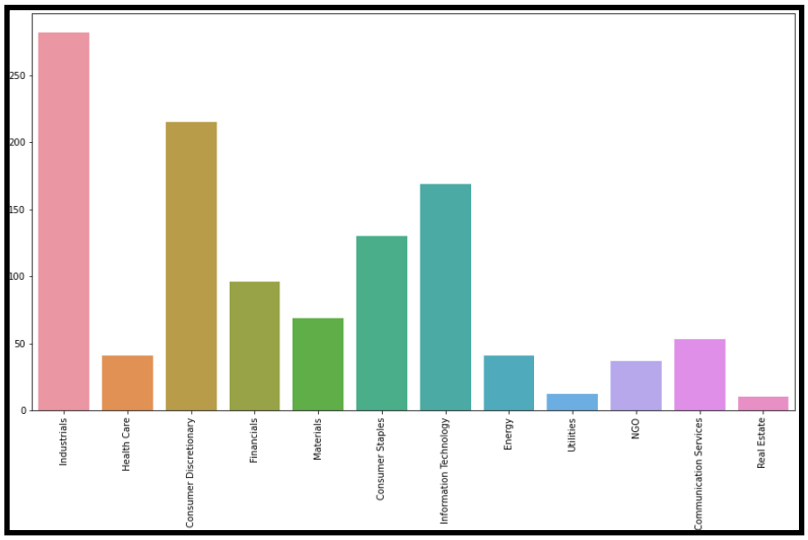
“ Yale CELI List” is a list of companies that left Russia or imposed sanctions against it or decided to stop their business in the country. 1156 unique companies can be found in the list grouped by their origin country in this way:



Countries Distribution

Most of the companies stopped their business in Russia or imposed the sanctions. Country of origin was not the main factor of such actions, however companies that focus on Information Technologies, Communication Services and Material Production tend to exit all business operations in Russia. Here is the distribution of companies based on their main industry focus:

As we can see, about 1/4 of all companies belong to industries mentioned above.



Industries Distribution

Son Ekaterina

Creation of a database of companies

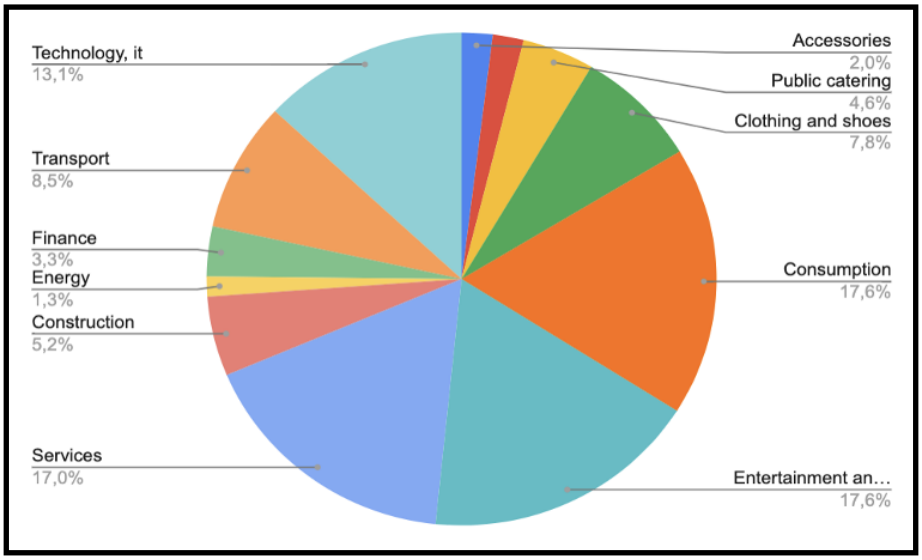
First of all, our task was to create a database with all companies that imposed limitations in response to sanctions or departed the Russian market in 2022. My part was to collect all data from "Sovcomblog", which is a blog about money management with articles on current topics, owned by Sovcombank. The website presents 282 companies that have left the Russian market and also analyzes in which area of the economy the most companies left. To collect all the data I copied names of companies that were represented on the blog website and pasted them in the common database. The purpose of this article is to summarize all the information about exiting companies. As It is challenging to pinpoint the precise number because media outlets practically daily report on corporate choices. The main source of data used for writing - "media". Based on the information given in the article, we can conclude that the companies in the field of "Clothing and retail" left the most.

Alexandra Arkhipova

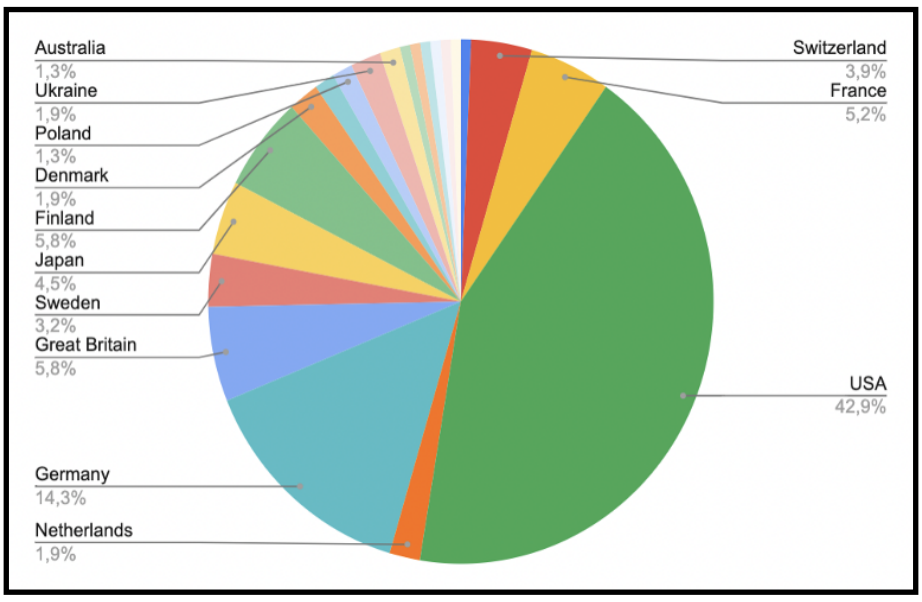
Was working with the web-page called **ROZETKED -** a website that publishes reviews and news from the world of IT and technology.

Link to the article: https://rozetked.me/cancellations/
Link to the list of companies collected from the article: List of companies

This website released a list of companies that stopped operating in Russia or put certain constraints on its market in February 2022. This article listed 354 businesses, the majority of which were American and the minority being Polish and Australian (chart 1). The analysis of the businesses mentioned in this report also revealed that the majority of them were in the consumer and entertainment industries, with services coming in second and technology & IT rounding out the top three (chart 2). However, the majority of the businesses have recently stopped their activities or deliveries in Russia.



Industries



Countries

Suchkov Ilya

Data:

I was working on the database "left-russia", which is originally is github's repository (https://github.com/daniilak/left-russia). Partially, the repository and database was compiled on the basis of information taken from the Telegram channel “Раньше всех. Ну почти” (https://t.me/bbbreaking). The list mostly includes European, US, Chinese and Australian companies that somehow reacted on of the situation around Ukraine. The last update of the database was 6.08.2022. Seven companies are placed in separate files with a more detailed description of actions after the start of sanctions.

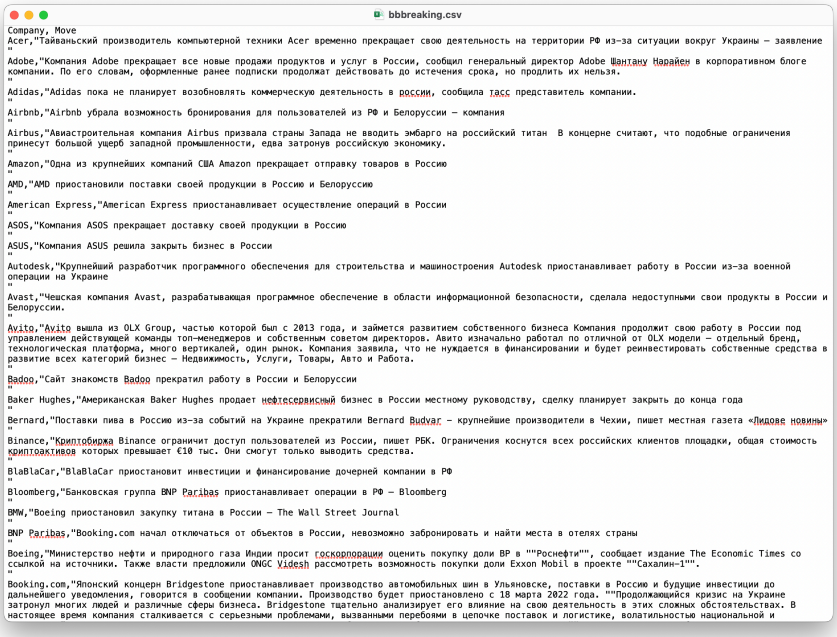
Working with data:

I wrote a 150-lines-code that parses data from a row ".md" file. There were few difficulties, since the author did not observe a clear number of indents, spaces, and so on. Nevertheless, without resorting for the most part of code (I used pandas at the end) to third-party libraries, I was able to transfer data to a file.

I can rate the work of the parser at 6 out of 10, since 8 lines out of 400 were translated incorrectly, which caused a slight shift in the table. I adjusted the shift manually. Given that the code coped with the task by 99.98 percent - we can say that it is not bad.

Result & Plans:

As a result, I got a ".csv" file that, actually, does not have any data of the country or industries, so I cannot show picturesque diagrams as my colleagues does. In the following two-three month I am going to use clustering analysis. To do this my data will be nested with industries and countries, so this is not a big deal. Furthermore, to acquaint you with the result a little closer, here's a little snippet of resulting ".csv" file:


bbbreaking.csv

```
Company, Move
Acer,"Тайваньский производитель компьютерной техники Acer временно прекращает свою деятельность на территории РФ из-за ситуации вокруг Украины — заявление
"
Adobe,"Компания Adobe прекращает все новые продажи продуктов и услуг в России, сообщил генеральный директор Adobe Шантану Нарайен в корпоративном блоге
компании. По его словам, оформленные ранее подписки продолжат действовать до истечения срока, но продлить их нельзя.
"
Adidas,"Adidas пока не планирует возобновлять коммерческую деятельность в россии, сообщила тасс представитель компании.
"
Airbnb,"Airbnb убрала возможность бронирования для пользователей из РФ и Белоруссии — компания
"
Airbus,"Авиастроительная компания Airbus призвала страны Запада не вводить эмбарго на российский титан  В концерне считают, что подобные ограничения
принесут большой ущерб западной промышленности, едва затронув российскую экономику.
"
Amazon,"Одна из крупнейших компаний США Amazon прекращает отправку товаров в Россию
"
AMD,"AMD приостановили поставки своей продукции в Россию и Белоруссию
"
American Express,"American Express приостанавливает осуществление операций в России
"
ASOS,"Компания ASOS прекращает доставку своей продукции в Россию
"
ASUS,"Компания ASUS решила закрыть бизнес в России
"
Autodesk,"Крупнейший разработчик программного обеспечения для строительства и машиностроения Autodesk приостанавливает работу в России из-за военной
операции на Украине
"
Avast,"Чешская компания Avast, разрабатывающая программное обеспечение в области информационной безопасности, сделала недоступными свои продукты в России и
Белоруссии.
"
Avito,"Avito вышла из OLX Group, частью которой был с 2013 года, и займется развитием собственного бизнеса Компания продолжит свою работу в России под
управлением действующей команды топ-менеджеров и собственным советом директоров. Авито изначально работал по отличной от OLX модели — отдельный бренд,
технологическая платформа, много вертикалей, один рынок. Компания заявила, что не нуждается в финансировании и будет реинвестировать собственные средства в
развитие всех категорий бизнес — Недвижимость, Услуги, Товары, Авто и Работа.
"
Badoo,"Сайт знакомств Badoo прекратил работу в России и Белоруссии
"
Baker Hughes,"Американская Baker Hughes продает нефтесервисный бизнес в России местному руководству, сделку планирует закрыть до конца года
"
Bernard,"Поставки пива в Россию из-за событий на Украине прекратили Bernard Budvar - крупнейшие производители в Чехии, пишет местная газета «Лидове новины»
"
Binance,"Криптобиржа Binance ограничит доступ пользователей из России, пишет РБК. Ограничения коснутся всех российских клиентов площадки, общая стоимость
криптоактивов которых превышает €10 тыс. Они смогут только выводить средства.
"
BlaBlaCar,"BlaBlaCar приостановит инвестиции и финансирование дочерней компании в РФ
"
Bloomberg,"Банковская группа BNP Paribas приостанавливает операции в РФ — Bloomberg
"
BMW,"Boeing приостановил закупку титана в России — The Wall Street Journal
"
BNP Paribas,"Booking.com начал отключаться от объектов в России, невозможно забронировать и найти места в отелях страны
"
Boeing,"Министерство нефти и природного газа Индии просит госкорпорации оценить покупку доли BP в ""Роснефти"", сообщает издание The Economic Times со
ссылкой на источники. Также власти предложили ONGC Videsh рассмотреть возможность покупки доли Exxon Mobil в проекте ""Сахалин-1"".
"
Booking.com,"Японский концерн Bridgestone приостанавливает производство автомобильных шин в Ульяновске, поставки в Россию и будущие инвестиции до
дальнейшего уведомления, говорится в сообщении компании. Производство будет приостановлено с 18 марта 2022 года. ""Продолжающийся кризис на Украине
затронул многих людей и различные сферы бизнеса. Bridgestone тщательно анализирует его влияние на свою деятельность в этих сложных обстоятельствах. В
настоящее время компания сталкивается с серьезными проблемами, вызванными перебоями в цепочке поставок и логистике, волатильностью национальной и
```

Next steps

As our next step we are going to split into two teams to work on different objectives. The first team is planning to gather more detailed information and useful characteristics of companies using clustering analysis. The second team is going to work on company sentiment analysis using neural networks.

Clustering analysis can be used to collect valuable insights about companies, such as their location, industry, size, and more. This data can be used to identify potential customer segments and guide alternative product development. Meanwhile, sentiment analysis with neural networks may be utilized to acquire information into client attitude and make accurate judgements regarding marketing tactics, product development, and service and support.

Group 1 (clustering analysis) — Arkhipova Alexandra, Suchkov Ilya, Lyakhova Daria.

We are using clustering analysis to get more familiar with data and make some assumption as the result. Many people know that cluster analysis is a "sub-chapter" of unsupervised machine learning. Therefore, this analysis will not be accomplished without the use of machine learning. At the end, this analysis will be applied using cluster algorithms, which give us a more classical and more fundamental data representation than sentiment analysis with RNN. So, we will get the full range of metrics from both analysis, that are needed for final conclusions and assumptions.

Group 2 (sentiment analysis with neural networks) — Melissa Dibé, Oganesyan Tigran, Son Ekaterina, Kruglikov Gleb.

We are going to use Recurrent Neural Network (RNN) for sentiment analysis. RNNs are a sort of artificial neural network that can learn from data sequences. That is, they may learn to recognise patterns in data and apply them to create predictions. This is beneficial for tasks like natural language processing, where word connections are vital. Using deep learning, it can identify the sentiment of a single sentence, a paragraph, or an entire text. Moreover, RNNs may learn the associations between words in a text and use them to categorise it as positive, negative, or neutral.