Indian Trade Data Analysis

By

Akarsh Somani (162) & Gaurav Misra (172)

Bachelor Thesis submitted to

Indian Institute of Information Technology Kalyani

For the partial fulfillment of the degree of

Bachelor of Technology

in

Computer Science and Engineering

April, 2019

Certificate

This is to certify that the thesis entitled "Indian Trade Data Analysis" being submitted by Akarsh Somani and Gaurav Misra, an undergraduate student, Reg. No 162 and 172, Roll No. 39/CSE/16005 and 39/CSE/16015, in the Department of Computer Science and Engineering, Indian Institute of Information Technology Kalyani, West Bengal 741235, India, for the award of Bachelors of Technology in Computer Science and Engineering is an original research work carried by him under my supervision and guidance. The thesis has fulfilled all the requirements as per the regulations of Indian Institute of Information Technology Kalyani and in my opinion, has reached the standards needed for submission. The work, techniques and the results presented have not been submitted to any other University or Institute for the award of any other degree or diploma.

Dalia Nandi

Assistant Professor (IIIT Kalyani)

Electronics and Communication Engineering Department

Declaration

I hereby declare that the work being presented in this thesis entitled, “Indian Trade Data Analysis”, submitted to Indian Institute of Information Technology Kalyani in partial fulfillment for the award of the degree of **Bachelor of Technology** in Computer Science and Engineering during the period from July, 2019 to November, 2019 under the supervision of Dr. Dalia Nandi, Department of Electronics and Communication Engineering, Indian Institute of Information Technology Kalyani, West Bengal 741235, India, does not contain any classified information.

Akarsh Somani and Gaurav Misra

Reg No/Roll No: 39/CSE/16005/162 and 39/CSE/16015/172

Department: Computer Science and Engineering

Institute Name: Indian Institute of Information Technology Kalyani

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

....

Dr. Dalia Nandi

Position: Assistant Professor

Departmental address: Electronics and Communication Engineering

Place: Kalyani

Date: 4/12/2019

Acknowledgments

Firstly, we would like to thank our supervisor Dr. Dalia Nandi for her support and guidance to complete this project work without whom we would not be able to make out this far. We would also like to thank our friends who supported us greatly and were always willing to help us. We are very grateful to Department of Computer Science and Engineering, Indian Institute of Information Technology Kalyani, West Bengal 741235, India, for providing us this wonderful opportunity. Lastly, we would like to thank our Parents a God for their never-ending grace.

Akarsh Somani and Gaurav Misra

Reg No/Roll No: 39/CSE/16005/162 and 39/CSE/16015/172

Department: Computer Science and Engineering

Institute Name: Indian Institute of Information Technology Kalyani

Place: Kalyani

Date: 4/12/2019

Abstract

This thesis investigates about the Indian Trade. There is a lot of things which is very important of the economy of our country and trade is one of them. We have done Data Analysis on the Indian trade and based on the results we will be drawing some conclusions.

Trade is the economic concept which invokes on BUY and Sell of the commodities, or exchanging goods and services. Trade increases competition and decreases overall world wise cost for a product. Here we will be dealing with the Indian trade with the other countries and the impact of the trade. Trade is one of the important factors for the economy of our country and that is why this trade data analysis is very important. All the commodities are divided into 99 Classes called HS Code (Harmonized System Code). Each class represent some commodity of trade and here our analysis starts. We have all the countries present from which India trade and for each country we have the HS Code which India trade from that particular country.

Looking at the above importance we tried to find some answers like

1. What HS Code we import/export the most?
2. From which country we import export the most?
3. Rate of change of trades in years.
4. From the country we import/export the most, what are the products HS Code traded.
5. Any future predictions on the Indian trade.
6. What category we trade the most year wise.
7. Highest trading country year wise.

Chapters

About Datasets (Yearly)

The data is taken from Kaggle - Link

The data set has two CSV file i.e. import data as well as export data from year 2010 to 2018.

Each data set has 5 columns -

1. HS Code – Harmonized System Code
2. Commodity – Description of the Harmonized System.
3. Value - Trade values for export and import of commodities in million US$
4. Country – Country from which import or export took place
5. Year – Year of import or export

HSCode	Commodity	value	country	year	type
5	PRODUCTS OF ANIMAL ORIGIN, NOT ELSEWHERE SPECIFIED OR INCLUDED.	0	AFGHANISTAN TIS	2018	import
7	EDIBLE VEGETABLES AND CERTAIN ROOTS AND TUBERS.	12.38	AFGHANISTAN TIS	2018	import
8	EDIBLE FRUIT AND NUTS; PEEL OR CITRUS FRUIT OR MELONS.	268.6	AFGHANISTAN TIS	2018	import
9	COFFEE, TEA, MATE AND SPICES.	35.48	AFGHANISTAN TIS	2018	import
11	PRODUCTS OF THE MILLING INDUSTRY; MALT; STARCHES; INULIN; WHEAT GLUTEN.		AFGHANISTAN TIS	2018	import
12	OIL SEEDS AND OLEA. FRUITS; MISC. GRAINS, SEEDS AND FRUIT; INDUSTRIAL OR MEDICINAL PLANTS	8.32	AFGHANISTAN TIS	2018	import
13	LAC; GUMS, RESINS AND OTHER VEGETABLE SAPS AND EXTRACTS.	108.78	AFGHANISTAN TIS	2018	import
20	PREPARATIONS OF VEGETABLES, FRUIT, NUTS OR OTHER PARTS OF PLANTS.	0.65	AFGHANISTAN TIS	2018	import
25	SALT; SULPHUR; EARTHS AND STONE; PLASTERING MATERIALS, LIME AND CEMENT.	0.05	AFGHANISTAN TIS	2018	import
27	MINERAL FUELS, MINERAL OILS AND PRODUCTS OF THEIR DISTILLATION; BITUMINOUS SUBSTANCES	0	AFGHANISTAN TIS	2018	import

HS Code - Link to Wikipedia

The **Harmonized Commodity Description and Coding System**, also known as the **Harmonized System** (**HS**) of tariff nomenclature is an internationally standardized system of names and numbers to classify traded products. It came into effect in 1988 and has since been developed and maintained by the World Customs Organization (WCO) (formerly the Customs Co-operation Council), an independent intergovernmental organization based in Brussels, Belgium, with over 200 member countries

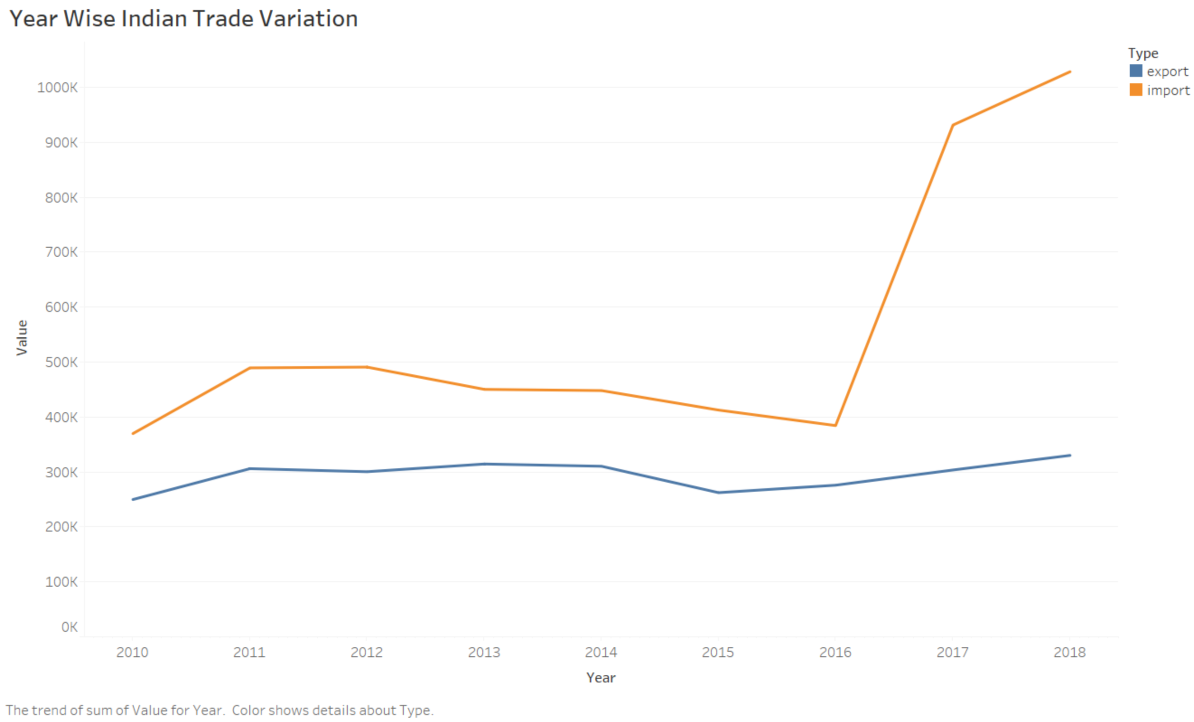
There are total 99 HS Code which consist of some commodity -

HS Code and their Description

HS Code	
1	LIVE ANIMALS.
2	MEAT AND EDIBLE MEAT OFFAL.
3	FISH AND CRUSTACEANS, MOLLUSCS AND OTHER AQUATIC INVERTABRATES.
4	DAIRY PRODUCE; BIRDS' EGGS; NATURAL HONEY; EDIBLE PROD. OF ANIMAL ORIGIN, NOT ELSEWHERE SPEC. OR INCLUDED.
5	PRODUCTS OF ANIMAL ORIGIN, NOT ELSEWHERE SPECIFIED OR INCLUDED.
6	LIVE TREES AND OTHER PLANTS; BULBS; ROOTS AND THE LIKE; CUT FLOWERS AND ORNAMENTAL FOLIAGE.
7	EDIBLE VEGETABLES AND CERTAIN ROOTS AND TUBERS.
8	EDIBLE FRUIT AND NUTS; PEEL OR CITRUS FRUIT OR MELONS.
9	COFFEE, TEA, MATE AND SPICES.
10	CEREALS.
11	PRODUCTS OF THE MILLING INDUSTRY; MALT; STARCHES; INULIN; WHEAT GLUTEN.
12	OIL SEEDS AND OLEA. FRUITS; MISC. GRAINS, SEEDS AND FRUIT; INDUSTRIAL OR MEDICINAL PLANTS; STRAW AND FODDER.
13	LAC; GUMS, RESINS AND OTHER VEGETABLE SAPS AND EXTRACTS.
14	VEGETABLE PLAITING MATERIALS; VEGETABLE PRODUCTS NOT ELSEWHERE SPECIFIED OR INCLUDED.
15	ANIMAL OR VEGETABLE FATS AND OILS AND THEIR CLEAVAGE PRODUCTS; PRE. EDIBLE FATS; ANIMAL OR VEGETABLE WAXEX.
16	PREPARATIONS OF MEAT, OF FISH OR OF CRUSTACEANS, MOLLUSCS OR OTHER AQUATIC INVERTEBRATES
17	SUGARS AND SUGAR CONFECTIONERY.
18	COCOA AND COCOA PREPARATIONS.
19	PREPARATIONS OF CEREALS, FLOUR, STARCH OR MILK; PASTRYCOOKS PRODUCTS.
20	PREPARATIONS OF VEGETABLES, FRUIT, NUTS OR OTHER PARTS OF PLANTS.
21	MISCELLANEOUS EDIBLE PREPARATIONS.
22	BEVERAGES, SPIRITS AND VINEGAR.
23	RESIDUES AND WASTE FROM THE FOOD INDUSTRIES; PREPARED ANIMAL FODER.
24	TOBACCO AND MANUFACTURED TOBACCO SUBSTITUTES.
25	SALT; SULPHUR; EARTHS AND STONE; PLASTERING MATERIALS, LIME AND CEMENT.
26	ORES, SLAG AND ASH.
27	MINERAL FUELS, MINERAL OILS AND PRODUCTS OF THEIR DISTILLATION; BITUMINOUS SUBSTANCES; MINERAL WAXES.
28	INORGANIC CHEMICALS; ORGANIC OR INORGANIC COMPOUNDS OF PRECIOUS METALS, OF RARE-EARTH METALS, OR RADI. ELEM. OR OF ISOTOPES.
29	ORGANIC CHEMICALS
30	PHARMACEUTICAL PRODUCTS
31	FERTILISERS.
32	TANNING OR DYEING EXTRACTS; TANNINS AND THEIR DERI. DYES, PIGMENTS AND OTHER COLOURING MATTER; PAINTS AND VER; PUTTY AND OTHER MASTICS; INKS.
33	ESSENTIAL OILS AND RESINOIDS; PERFUMERY, COSMETIC OR TOILET PREPARATIONS.
34	SOAP, ORGANIC SURFACE-ACTIVE AGENTS, WASHING PREPARATIONS, LUBRICATING PREPARATIONS, ARTIFICIAL WAXES, PREPARED WAXES, POLISHING OR SCOURING PR..
35	ALBUMINOIDAL SUBSTANCES; MODIFIED STARCHES; GLUES; ENZYMES.
36	EXPLOSIVES; PYROTECHNIC PRODUCTS; MATCHES; PYROPHORIC ALLOYS; CERTAIN COMBUSTIBLE PREPARATIONS.
37	PHOTOGRAPHIC OR CINEMATOGRAPHIC GOODS.
38	MISCELLANEOUS CHEMICAL PRODUCTS.
39	PLASTIC AND ARTICLES THEREOF.
40	RUBBER AND ARTICLES THEREOF.
41	RAW HIDES AND SKINS (OTHER THAN FURSKINS) AND LEATHER
42	ARTICLES OF LEATHER,SADDLERY AND HARNESS;TRAVEL GOODS, HANDBAGS AND SIMILAR CONT.ARTICLES OF ANIMAL GUT(OTHR THN SILK-WRM)GUT.
43	FURSKINS AND ARTIFICIAL FUR, MANUFACTURES THEREOF.
44	WOOD AND ARTICLES OF WOOD; WOOD CHARCOAL.
45	CORK AND ARTICLES OF CORK.
46	MANUFACTURES OF STRAW, OF ESPARTO OR OF OTHER PLAITING MATERIALS; BASKETWARE AND WICKERWORK.
47	PULP OF WOOD OR OF OTHER FIBROUS CELLULOSIC MATERIAL; WASTE AND SCRAP OF PAPER OR PAPERBOARD.
48	PAPER AND PAPERBOARD; ARTICLES OF PAPER PULP, OF PAPER OR OF PAPERBOARD.
49	PRINTED BOOKDS, NEWSPAPERS, PICTURES AND OTHER PRODUCTS OF THE PRINTING INDUSTRY; MANUSCRIPTS, TYPESCRIPTS AND PLANS.
50	SILK
51	WOOL, FINE OR COARSE ANIMAL HAIR, HORSEHAIR YARN AND WOVEN FABRIC.
52	COTTON.
53	OTHER VEGETABLE TEXTILE FIBRES; PAPER YARN AND WOVEN FABRICS OF PAPER YARN.
54	MAN-MADE FILAMENTS.
55	MAN-MADE STAPLE FIBRES.
56	WADDING, FELT AND NONWOVENS; SPACIAL YARNS; TWINE, CORDAGE, ROPES AND CABLES AND ARTICLES THEREOF.
57	CARPETS AND OTHER TEXTILE FLOOR COVERINGS.
58	SPECIAL WOVEN FABRICS; TUFTED TEXTILE FABRICS; LACE; TAPESTRIES; TRIMMINGS; EMBROIDERY.
59	IMPREGNATED, COATED, COVERED OR LAMINATED TEXTILE FABRICS; TEXTILE ARTICLES OF A KIND SUITABLE FOR INDUSTRIAL USE.
60	KNITTED OR CROCHETED FABRICS.
61	ARTICLES OF APPAREL AND CLOTHING ACCESSORIES, KNITTED OR CORCHETED.
62	ARTICLES OF APPAREL AND CLOTHING ACCESSORIES, NOT KNITTED OR CROCHETED.
63	OTHER MADE UP TEXTILE ARTICLES; SETS; WORN CLOTHING AND WORN TEXTILE ARTICLES; RAGS
64	FOOTWEAR, GAITERS AND THE LIKE; PARTS OF SUCH ARTICLES.
65	HEADGEAR AND PARTS THEREOF.
66	UMBRELLAS, SUN UMBRELLAS, WALKING-STICKS, SEAT-STICKS, WHIPS,RIDING-CROPS AND PARTS THEREOF.
67	PREPARED FEATHERS AND DOWN AND ARTICLES MADE OF FEATHERS OR OF DOWN; ARTIFICIAL FLOWERS; ARTICLES OF HUMAN HAIR.
68	ARTICLES OF STONE, PLASTER, CEMENT, ASBESTOS, MICA OR SIMILAR MATERIALS.
69	CERAMIC PRODUCTS.
70	GLASS AND GLASSWARE.
71	NATURAL OR CULTURED PEARLS,PRECIOUS OR SEMIPRECIOUS STONES,PRE.METALS,CLAD WITH PRE.METAL AND ARTCLS THEREOF;IMIT.JEWLRY;COIN.
72	IRON AND STEEL
73	ARTICLES OF IRON OR STEEL
74	COPPER AND ARTICLES THEREOF.
75	NICKEL AND ARTICLES THEREOF.
76	ALUMINIUM AND ARTICLES THEREOF.
78	LEAD AND ARTICLES THEREOF.
79	ZINC AND ARTICLES THEREOF.
80	TIN AND ARTICLES THEREOF.
81	OTHER BASE METALS; CERMETS; ARTICLES THEREOF.
82	TOOLS IMPLEMENTS, CUTLERY, SPOONS AND FORKS, OF BASE METAL; PARTS THEREOF OF BASE METAL.
83	MISCELLANEOUS ARTICLES OF BASE METAL.
84	NUCLEAR REACTORS, BOILERS, MACHINERY AND MECHANICAL APPLIANCES; PARTS THEREOF.
85	ELECTRICAL MACHINERY AND EQUIPMENT AND PARTS THEREOF; SOUND RECORDERS AND REPRODUCERS, TELEVISION IMAGE AND SOUND RECORDERS AND REPRODUCE..
86	RAILWAY OR TRAMWAY LOCOMOTIVES, ROLLING-STOCK AND PARTS THEREOF; RAILWAY OR TRAMWAY TRACK FIXTURES AND FITTINGS AND PARTS THEREOF; MECHANICAL
87	VEHICLES OTHER THAN RAILWAY OR TRAMWAY ROLLING STOCK, AND PARTS AND ACCESSORIES THEREOF.
88	AIRCRAFT, SPACECRAFT, AND PARTS THEREOF.
89	SHIPS, BOATS AND FLOATING STRUCTURES.
90	OPTICAL, PHOTOGRAPHIC CINEMATOGRAPHIC MEASURING, CHECKING PRECISION, MEDICAL OR SURGICAL INST. AND APPARATUS PARTS AND ACCESSORIES THEREOF;
91	CLOCKS AND WATCHES AND PARTS THEREOF.
92	MUSICAL INSTRUMENTS; PARTS AND ACCESSORIES OF SUCH ARTICLES.
93	ARMS AND AMMUNITION; PARTS AND ACCESSORIES THEREOF.
94	FURNITURE; BEDDING, MATTRESSES, MATTRESS SUPPORTS, CUSHIONS AND SIMILAR STUFFED FURNISHING; LAMPS AND LIGHTING FITTINGS NOT ELSEWHERE SPECIFIED ..
95	TOYS, GAMES AND SPORTS REQUISITES; PARTS AND ACCESSORIES THEREOF.
96	MISCELLANEOUS MANUFACTURED ARTICLES.
97	WORKS OF ART COLLECTORS' PIECES AND ANTIQUES.
98	PROJECT GOODS; SOME SPECIAL USES.
99	MISCELLANEOUS GOODS.

Commodity broken down by HS Code.

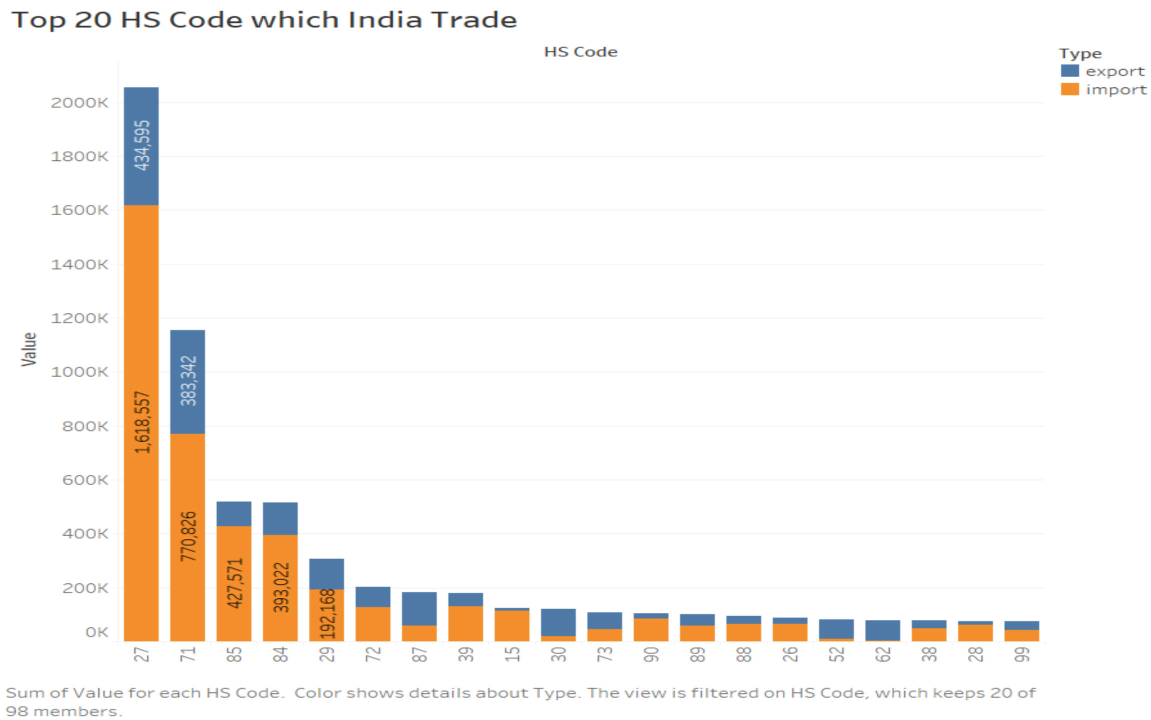
Year Wise Trade



We can see a bump in the year 2017 and 2018. It can be because of the fact that GST came in picture that year and thus the custom duty was revised.

This huge increase in import can also be a factor of the decreasing Indian economy.

HS Code wise Trade



27 - MINERAL FUELS, MINERAL OILS AND PRODUCTS OF THEIR DISTILLATION; BITUMINOUS SUBSTANCES; MINERAL WAXES

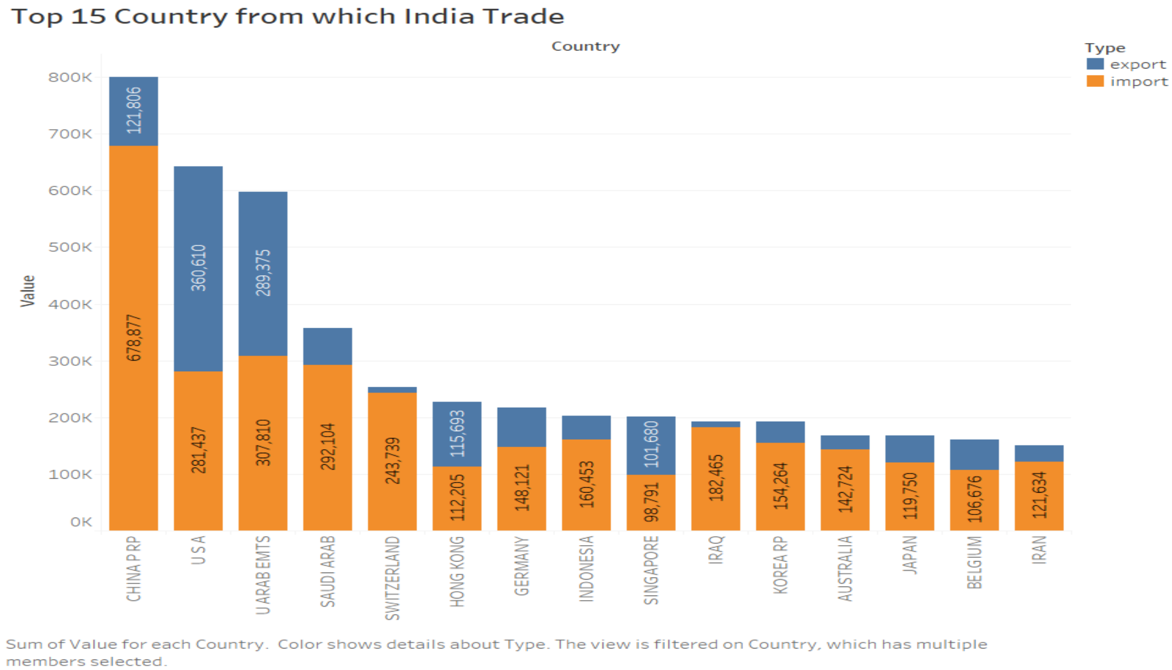
71 - NATURAL OR CULTURED PEARLS, PRECIOUS OR SEMIPRECIOUS STONES, PRE.METALS, CLAD WITH PRE.METAL AND ARTCLS THEREOF; IMIT.JEWLRY; COIN.

85 - ELECTRICAL MACHINERY AND EQUIPMENT AND PARTS THEREOF; SOUND RECORDERS AND REPRODUCERS, TELEVISION IMAGE AND SOUND RECORDERS AND REPRODUCERS, AND PARTS.

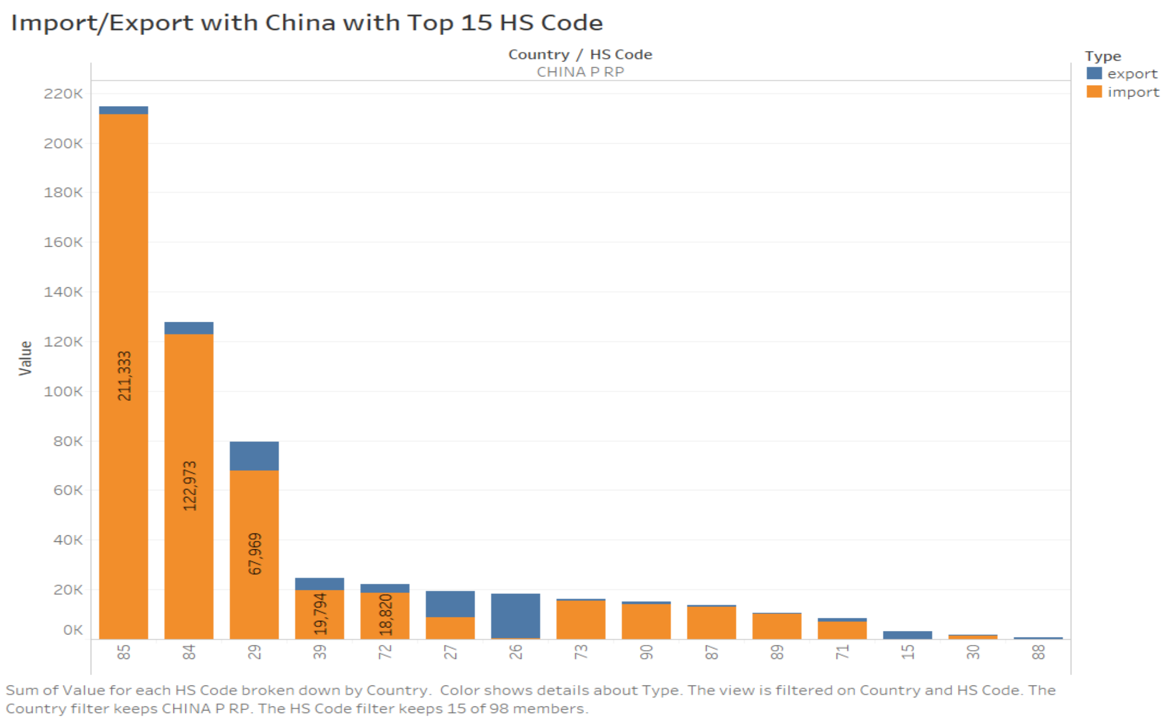
84 - NUCLEAR REACTORS, BOILERS, MACHINERY AND MECHANICAL APPLIANCES; PARTS THEREOF.

And so on…

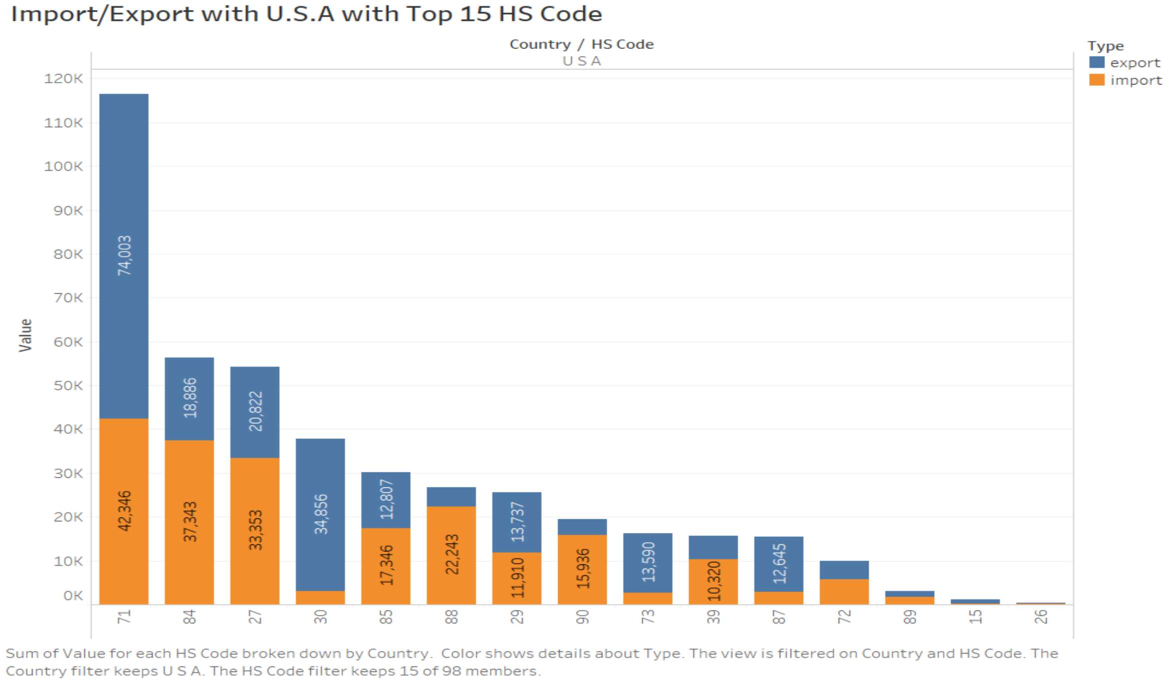
Country wise Trade



Maximum import is from China and the Maximum export is to U.S.A.



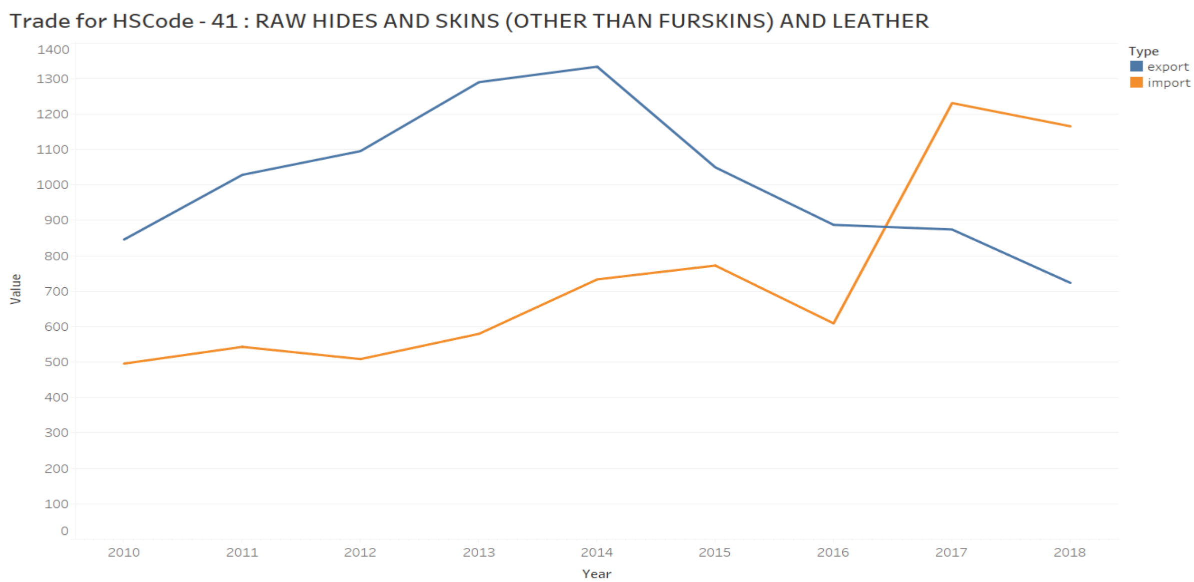
84, 85 are Machinery equipment which is justified India is a big importer of machineries from the china.



71 is pearls and precious stones whereas 84 is machinery and 27 is natural oils.

Inferences on some HS Codes

HS Code- 41(Hides, Skins and Leathers)

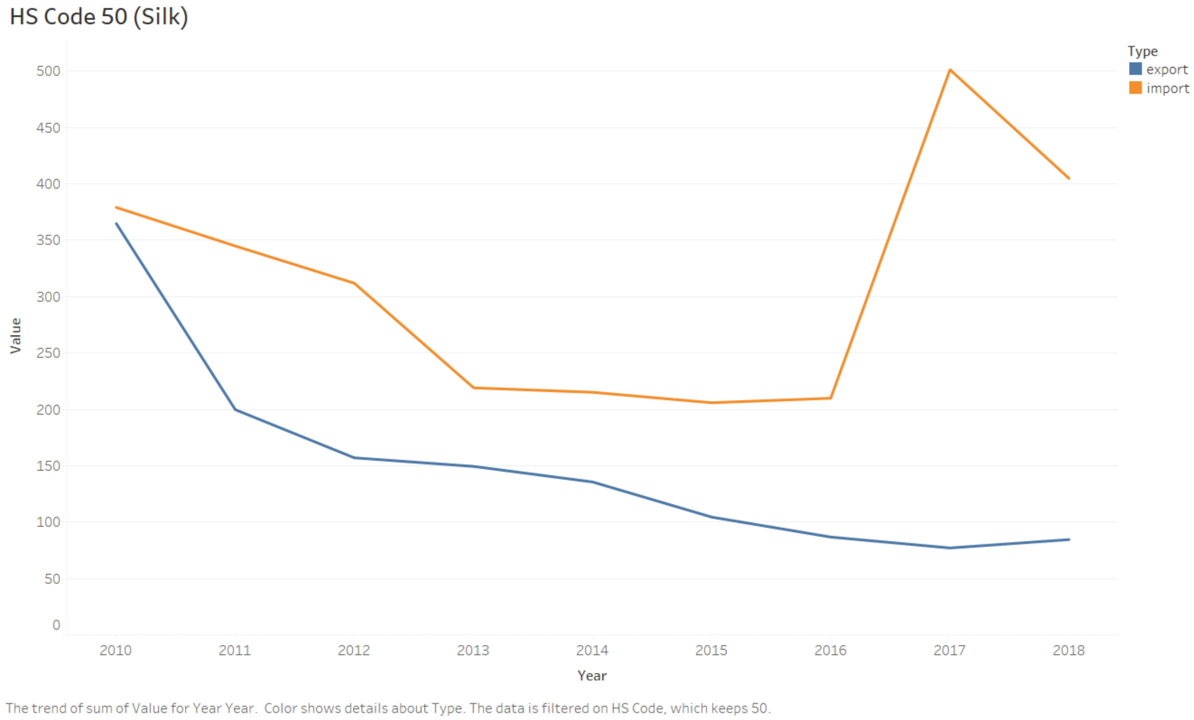


The trend of sum of Value for Year Year. Color shows details about Type. The data is filtered on HS Code, which keeps 41.

India was a huge exporter of the leather, hides and skins but the controversies and violence of killing animals shut down several factories which lead to decrease in export of leather, skins and hides while import increases.

“Export growth slowed to 9.37% in 2014-15 and declined more than 19 percentage points in 2015-16, the data show.” - by Indiaspend[1]

HS Code- 50 (Silk Products)



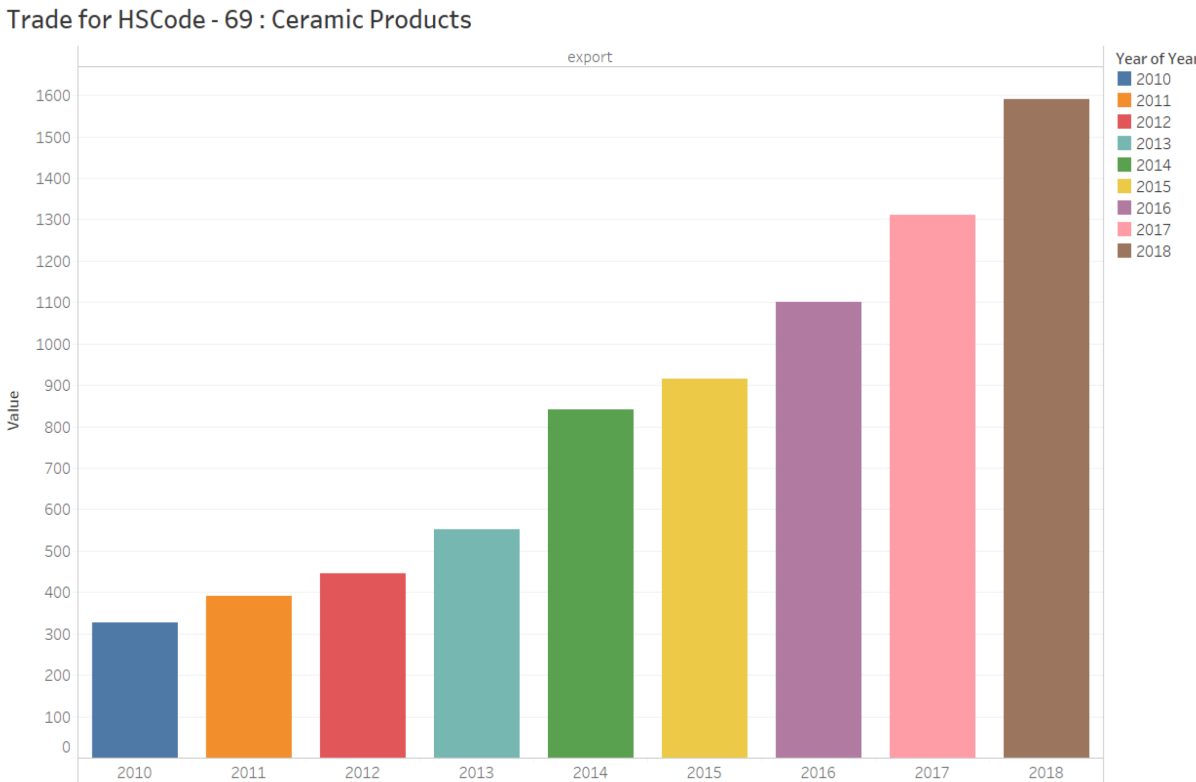
As we can see that the export of the silk is decreasing and we tried to find the reasons behind it.

Vietnam a major threat -

Currently, any import of silk fabrics from Vietnam attracts zero per cent duty, which makes India's silk fabrics expensive. This has significantly affected the domestic market of India.

Bharat Gandhi, Chairman, The Federation of Indian Art Silk Weaving Industry, averred that the phenomenal rise in the import of silk fabrics from Vietnam in last 2 years has deteriorated the conditions of silk fabric manufacturers hailing from Indian cities of Bhagalpur, Varanasi, Bengaluru, Surat and some parts of Tamil Nadu.[2]

HS Code – 69 (Ceramic product)



Sum of Value for each Year Year broken down by Type. Color shows details about Year Year. The data is filtered on HS Code, which keeps 69. The view is filtered on Type, which keeps export.

It is good to see that the Indian export of ceramic products are constantly increasing.

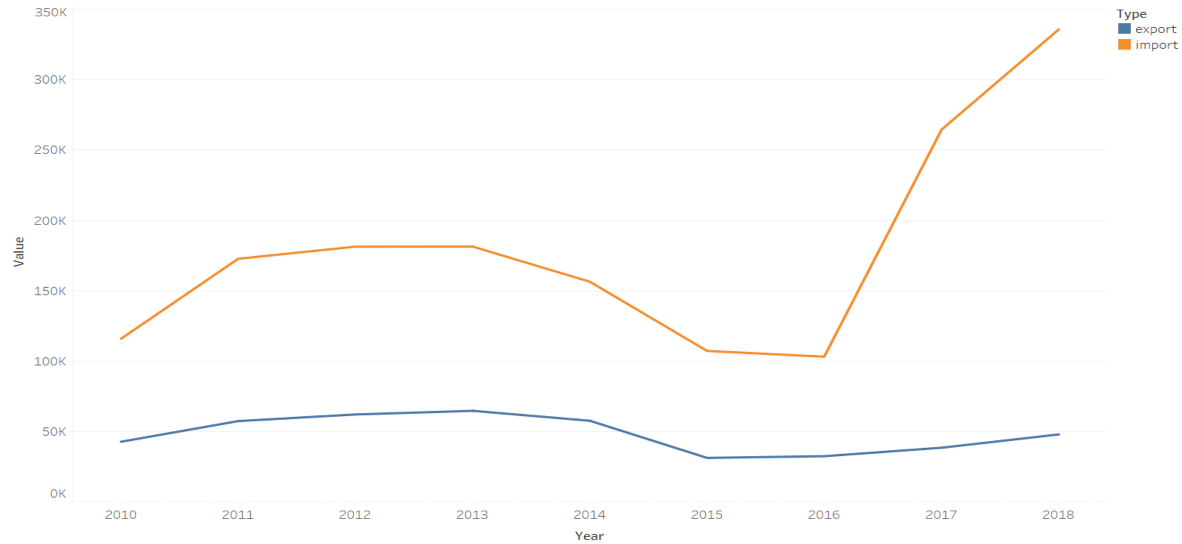
"During 2017, India was the 24th largest ceramic trading nation in the world and accounted for a share of around 0.9% in total ceramics trade. During the period, from 2010 to 2017, India's ceramics trade increased from US $836.8 million to US $1.8 billion at a CAGR of 10.07%."

-By CATR, (Centre for Advance Trade Research) | August 29, 2018[3]

HS Code – 27

(MINERAL FUELS, MINERAL OILS AND PRODUCTS OF THEIR DISTILLATION; BITUMINOUS SUBSTANCES; MINERAL WAXES.)

HS Code - 27 : Year wise Import/ Export

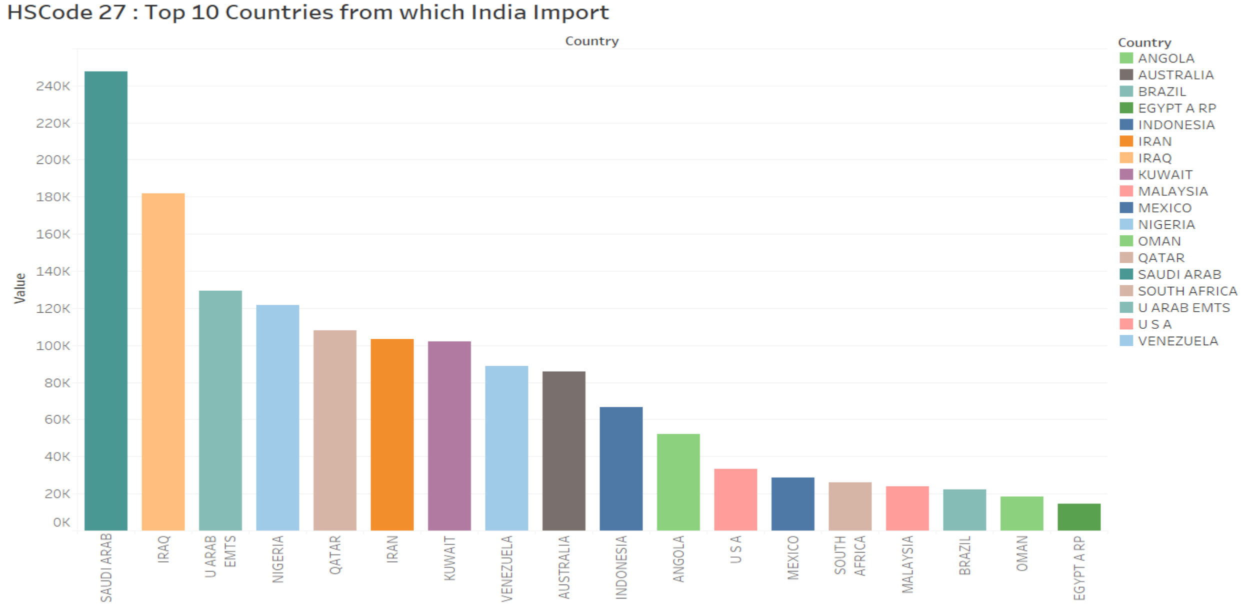


The trend of sum of Value for Year Year. Color shows details about Type. The data is filtered on HS Code, which keeps 27.

Import is increased by huge amount after the custom duty revision from 2017.

"Petroleum and crude oil imports continued to inflate India's import bill, rising 42.64% from last year to $11.65 billion." -Feb 16, 2018, India Times [4]

HSCode 27 : Top 10 Countries from which India Import



Sum of Value for each Country. Color shows details about Country. The data is filtered on HS Code and Type. The HS Code filter keeps 27. The Type filter keeps import. The view is filtered on Country, which keeps 18 members.

Saudi Arab is the highest followed by IRAQ, United Arab Emirates, Nigeria, Qatar, IRAN and Kuwait are some major countries from which India import Minerals products.

Forecasting on Monthly Trade Data

Significance of Forecasting

What if we can predict the next month's total import or export amount? The question depends on how much accuracy can be achieved while forecasting future data points. Assuming we get fairly well accuracy, we can surely say that there's a lot of use cases to get returns with forecasting total trade amount.

We purchase things regularly out of our need. Sometimes more and sometimes less, but if we know that next month's import is going to be lower, we can say that prices are also going to get cheaper. The reason behind lowering of price is that either demand is slow or we are producing that product on a larger scale in our country, hence the decrement in import.

The most specific use-case appear to be for industries, since they buy very precious machineries for functionality and production purpose, hence if they can somehow know that imports are going to decrease and hence prices of machines can be cheaper they can plan to buy machines earlier than usual and same for delaying the purchase in case of increment in imports and decrement in exports.

Government can plan to handle demands of next month using forecasted imported value similarly for exports as well. Government can also control prices based on whether imports and exports are going to go up or down.

Data (Monthly)

Dataset has been scraped from Department of Commerce, Govt. of India. Data is available from January, 2006 to September, 2019. We have total trade amount (import/export) for each month which lies in ranges in million US dollars.

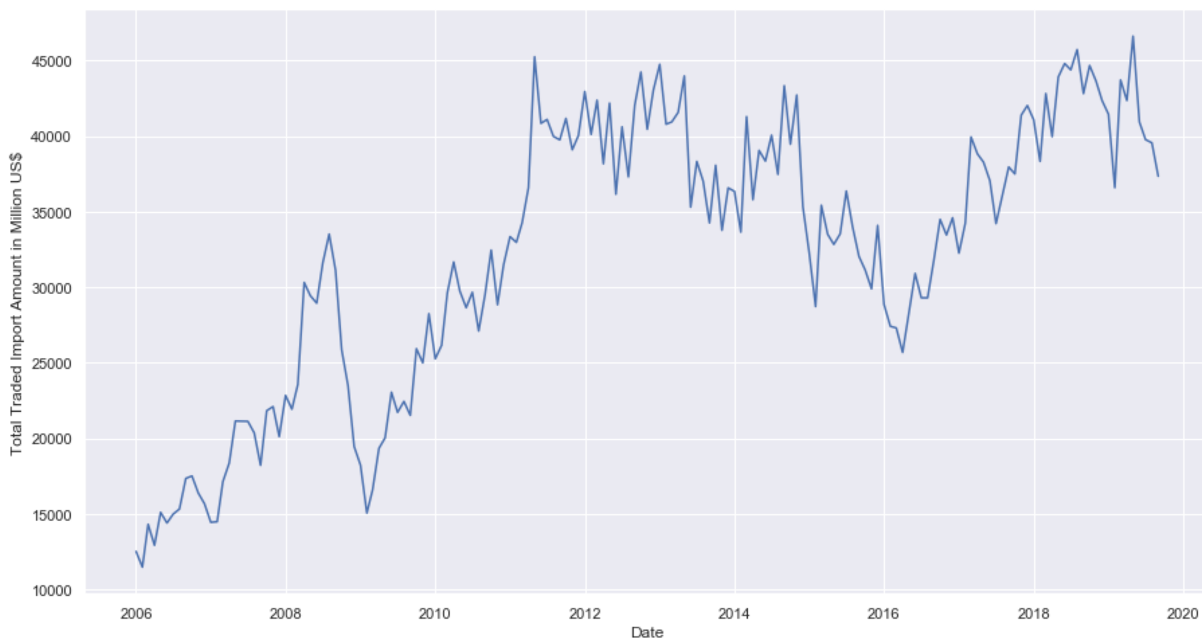
Date	Import	Export
Jan-06	12,519.71	9,143.66
Feb-06	11,479.69	8,993.29
Mar-06	14,314.02	11,560.97
Apr-06	12,924.18	8,624.66
May-06	15,105.67	10,109.30
Jun-06	14,399.73	10,419.60
Jul-06	14,985.01	10,599.73
Aug-06	15,326.36	10,769.09
Sep-06	17,350.69	10,756.49

Preprocessing of Dataset

We have total 165 data points. Now we break the dataset into two sets – training and validation such that training set includes 150 data points and validation set includes 15 data points to forecast upon.

We have no missing values hence we don't need to impute any data point. We have all numerical columns hence we don't need to one hot encode as well.

Visualization



Forecasting Models

- **Moving Average Model:** In time series analysis, the Moving-Average model (MA model), also known as moving-average process, is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.
- **ARIMA Model:** Auto Regressive Integrated Moving Average (ARIMA) models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity.
- **CNN Model:** In deep learning, a Convolutional Neural Network (CNN) is a class of deep neural networks, most commonly applied to analyzing visual imagery.
- **RNN Model:** A Recurrent Neural Network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior.
- **LSTM Model:** Long Short-Term Memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video).

Which one to use

- Moving Average models may perform moderately in short term forecasting but are not reliable since there's no weightage of past values taken.
- ARIMA method is appropriate only for a time series that is stationary.
- CNNs are inappropriate for sequential data as they can't use past values to make predictions.
- RNNs are used for processing sequential data but perform badly in case of long-term dependency, this happens because of vanishing gradient problem.
- LSTMs store past information and know what to forget and what not to forget as a modification over RNNs.

"Hence LSTM models are best to process and forecast sequential data specially time series data. They have powerful memory to store past values."

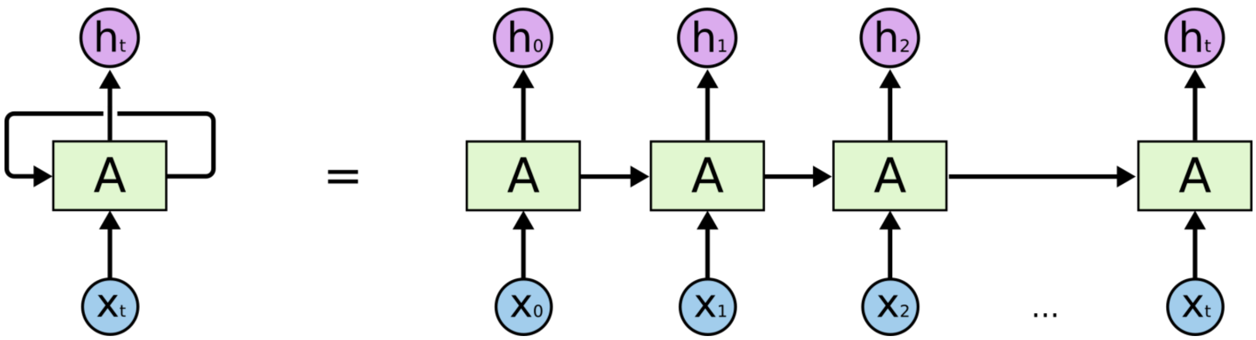
Before understanding LSTM models, we need to understand RNN models.

RNN

Humans don't start their thinking from scratch every second. As we read this essay, we understand each word based on your understanding of previous words. We don't throw everything away and start thinking from scratch again. Our thoughts have persistence.

Traditional neural networks can't do this, and it seems like a major shortcoming.

Recurrent neural networks address this issue. They are networks with loops in them, allowing information to persist.



Sometimes, we only need to look at recent information to perform the present task. In such cases, where the gap between the relevant information and the place that it's needed is small, RNNs can learn to use the past information.

But there are also cases where we need more contexts. Consider trying to predict the last word in the text "I grew up in France… I speak fluent French." It's entirely possible for the gap between the relevant information and the point where it is needed to become very large. Unfortunately, as that gap grows, RNNs become unable to learn to connect the information.

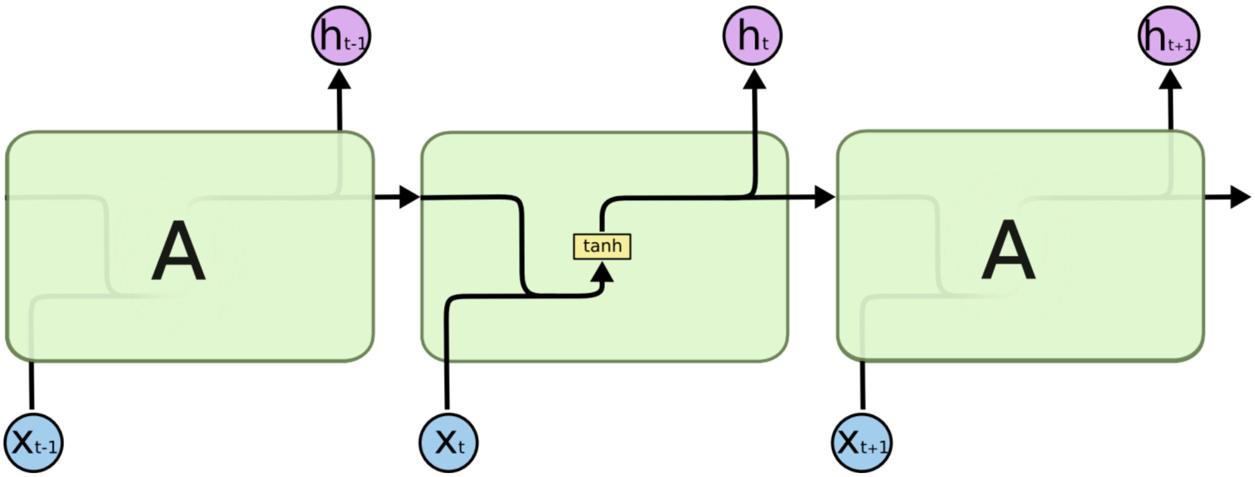
Thankfully, LSTMs Solves this problem!

LSTM Networks

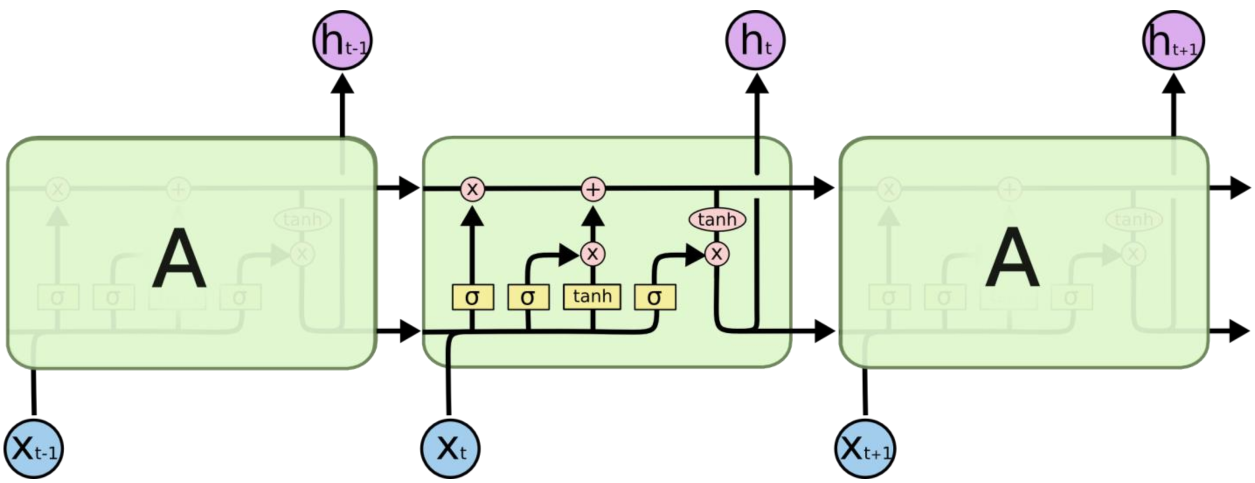
Long Short Term Memory networks – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter & Schmidhuber (1997), and were refined and popularized by many people in following work.

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.

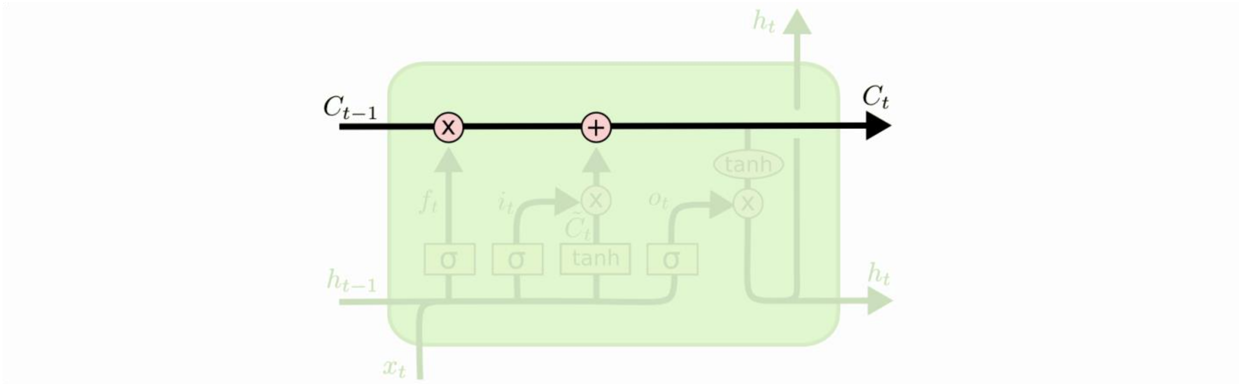


LSTMs also have this chain like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.



The key to LSTMs is the cell state, the horizontal line running through the top of the diagram.

The cell state is kind of like a conveyor belt. It runs straight down the entire chain, with only some minor linear interactions. It's very easy for information to just flow along it unchanged.



The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates.

Gates are a way to optionally let information through. They are composed out of a sigmoid neural net layer and a pointwise multiplication operation.

Preparing Data for applying LSTM

Scale data values between 0 and 1 for faster convergence using gradient descent. Prepare input for each of data point as set of previous 12 data point, e.g. for 13th data point a_{13} the input will be set (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}) and the output will be a_{13}. Store all these input sets in a input list and outputs in output list.

LSTM Model Specification

We have four layers model, which is arranged as following:

1. LSTM layer of 200 nodes with 0.2 dropout
2. LSTM layer of 200 nodes with 0.2 dropout
3. LSTM layer of 150 nodes with 0.2 dropout
4. Dense layer with ‘linear’ activation.

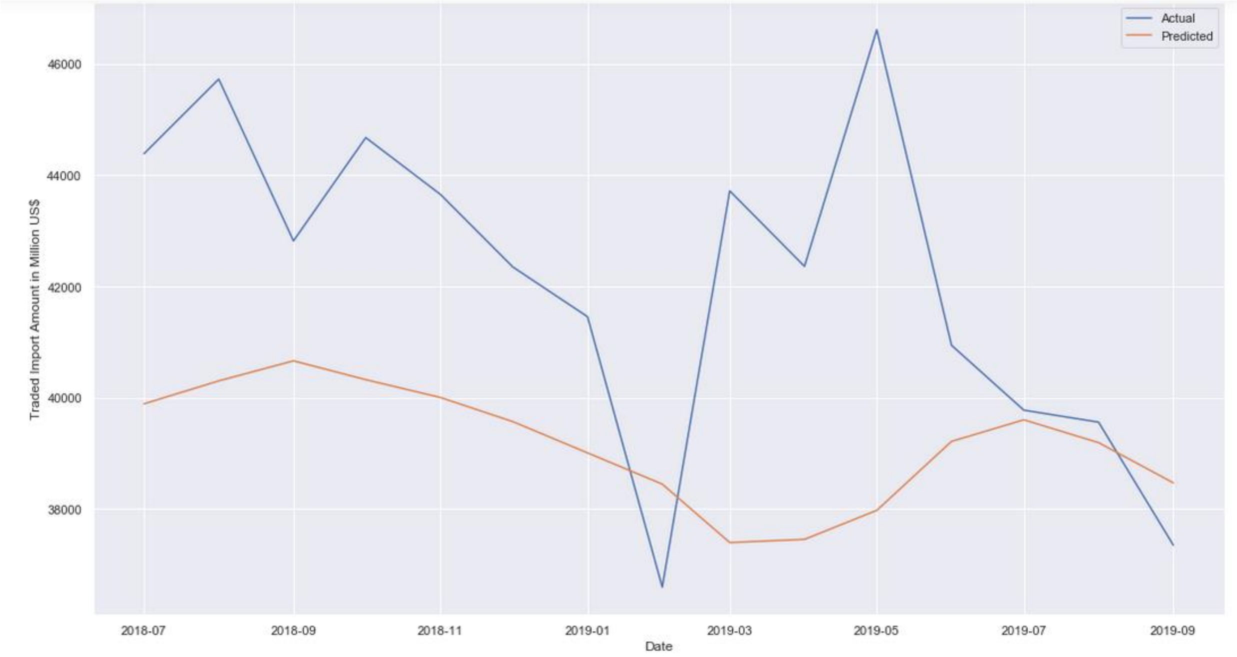
After defining our model, we compile it using mean-squared-error as loss function and Adam optimizer.

Now we fit our model for prepared input and output with batch size of 15 for 30 epochs.

Then we predict data values for next 15 months one by one, however we can do this in one go as well. These predicted values will be scaled in between 0 and 1 hence we need to inverse transform these values to get actual value for trade of that month.

Results

We have predicted total 15 data points with mean of 3363.970323 and standard deviation of 2352.545462.



Conclusion

- The mean of Error in forecasting is 3363.970323 and standard deviation is 2352.545462.
- To get more accurate forecasting we should fine tune our model's parameter in a better way.
- We can also modify our model's network structure.
- The brighter side is that our model is able to forecast right trend of trade import amount.
- As we have only 165 data points (150 training + 15 validations), it's too much to ask for a very high accuracy.
- More data points will improve our model's accuracy.

Technologies Used

Python, Jupyter Notebook, Tableau

Future Prospects

- Going back to data and searching all kind of patterns.
- Finding any major country impacting our import/export.
- Drawing more insights about trade.
- Collecting more data points.
- Forecasting for trade with specific HS-code or country.
- Fine tune existing model.

References

- https://www.indiaspend.com/indias-leather-exports-decline-as-cow-related-violence-increases-99395/
- https://apparelresources.com/business-news/sourcing/increase-import-silk-fabrics-vietnam-big-threat-india/
- https://www.tpci.in/blogs/product-profile-ceramics/
- https://economictimes.indiatimes.com/news/economy/indicators/indias-exports-fall-in-january-so-do-imports/articleshow/62934403.cms?from=mdr&utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst
- https://retail.regionaldirectory.us/ceramic-studios-and-products/
- https://www.ar.hwaiat.com/Products.aspx?CID=249&&new022d261d-a6c1-479b-9c72-21e922357299
- https://www.irishtimes.com/news/environment/increasing-fossil-fuel-use-push-carbon-emissions-to-record-high-1.3721474
- https://colah.github.io/posts/2015-08-Understanding-LSTMs/
- Wikipedia