

U.S. COFFEE IMPORTS: THE COMPETITION FOR MARKET SHARE
DATA SCIENCE DAT7 STUDENT PROJECT
BY SETRA YAPPI

1. BACKGROUND AND HYPOTHESIS

As the United States do not produce coffee beans, over 1.5 billion kilograms of coffee was imported into the country in 2014. Green coffee (unroasted coffee beans) is a commodity that is price-sensitive and prone to supply fluctuations (based on weather changes). It is also a commodity which has been sold under various certification schemes to ensure sustainability in the agricultural processes involved. Such certification may increase demand as well as pricing.

The United States as a country is the largest coffee consuming market in the world, and is a global trendsetter. Top producing countries compete for market share in the US. The initial hypothesis is that high supply availability, low prices and availability of certified coffee help increase market share in a given year. Coffee certifications include 4C Association, Rainforest Alliance, UTZ, Fairtrade and Organic.

In international trade, suppliers and buyers build long-term relationships, often involving contracts whereby supply during several months or a year is committed to a specific buyer, and therefore relationships are not easily terminated. For this project, we will look at how changes in price and supply over 3 years leading to the year of analysis affect the market share.

2. DATASETS AND FEATURE SELECTION

The model for this project is built upon 2014 (full year) price and supply information, as well as percentage of coffee certified by the various certification schemes (2012), which are the key market drivers in the coffee industry. Below is the list of available datasets used in this study:

- US import data: <http://dataweb.usitc.gov/>

Sample data:

Country	Coffee, not roasted, decaffeinated	Coffee, not roasted, not decaffeinated	Coffee, roasted, decaffeinated	Coffee, roasted, not decaffeinated	Grand Total	% Not Roasted/Decaf.
Brazil	\$68,510,981	\$1,264,840,675	\$419,147	\$7,208,634	\$1,340,979,437	94%
Colombia	\$52,564,289	\$1,076,251,877	\$1,490,643	\$33,657,789	\$1,163,964,598	92%
Vietnam	\$52,329,910	\$441,414,279	\$86,872	\$4,788,681	\$498,619,742	89%
Canada	\$9,582,398	\$2,186,907	\$30,594,982	\$343,223,580	\$385,587,867	1%
Guatemala	\$5,141,731	\$354,691,587	\$490,388	\$122,547	\$360,446,253	98%
Indonesia	\$23,649,215	\$298,773,577	\$77,364	\$604,012	\$323,104,168	92%
Mexico	\$42,113,352	\$210,878,781	\$466,510	\$29,907,281	\$283,365,924	74%
Nicaragua	\$7,034,015	\$221,210,140	\$26,553	\$644,545	\$228,915,253	97%
Peru	\$25,163,375	\$183,968,949	\$18,415	\$104,134	\$209,254,873	88%
Honduras	\$28,259,150	\$170,103,533	\$282,204	\$196,565	\$198,841,452	86%

Country	Custom Value 2011	Custom Value 2014	Custom Value - Percent Change 2011 - 2014	Quantity 2011, kg	Quantity 2014, kg	Quantity - Percent Change 2011 - 2014
Brazil	1,865,120,940	1,264,840,675	-32.20%	393,741,672	409,421,505	4.00%
Colombia	1,173,972,285	1,076,251,877	-8.30%	188,282,604	257,476,975	36.80%
Vietnam	455,892,223	441,414,279	-3.20%	189,602,810	207,996,961	9.70%
Guatemala	559,194,584	354,691,587	-36.60%	92,026,385	82,352,006	-10.50%
Canada	345,897,689	343,223,580	-0.80%	31,420,919	33,724,487	7.30%
Indonesia	301,362,213	298,773,577	-0.90%	54,203,674	60,036,292	10.80%
Nicaragua	205,466,101	221,210,140	7.70%	37,007,669	54,972,180	48.50%
Mexico	389,963,605	210,878,781	-45.90%	67,705,463	53,061,929	-21.60%
Peru	351,227,700	183,968,949	-47.60%	60,396,628	47,292,694	-21.70%
Costa Rica	240,802,635	176,035,207	-26.90%	40,444,186	39,305,506	-2.80%

- National annual coffee production by country: <http://www.nationmaster.com/country-info/stats/Agriculture/Crops/Beans/Coffee/Coffee-production>
- China coffee production: http://europe.chinadaily.com.cn/business/2014-12/24/content_19153826.htm
- Global trade data (import, export, re-export, re-import) <http://comtrade.un.org/data/>

- Percentage of certified coffee by country: Standard-Compliant production as a percentage of total national production For 20 largest coffee producers, 2012
https://www.iisd.org/pdf/2014/ssi_2014_chapter_8.pdf

For the uniformity of the analysis, the category of imported coffee selected is Coffee, not roasted, not decaffeinated (HTS 09011100), where coffee is priced per kilogram. Roasted and/or decaffeinated coffee often costs higher, and thus is not included in this study. Non-roasted-or-decaffeinated coffee, in the form of coffee beans or ground coffee, dominates coffee imports; for top exporting countries selling to US buyers, between 86%-97% of the total coffee exports to USA.

Countries selected for this study are only those who are net exporters of coffee in the world. This is calculated based on the sum of coffee exports and re-exports (imported coffee sold to overseas buyers) subtracted by coffee imports. Only countries with positive net coffee exports are included in the model. Also, countries with zero exports to USA in 2014 are disregarded.

Features selected for each country based on the above are the following:

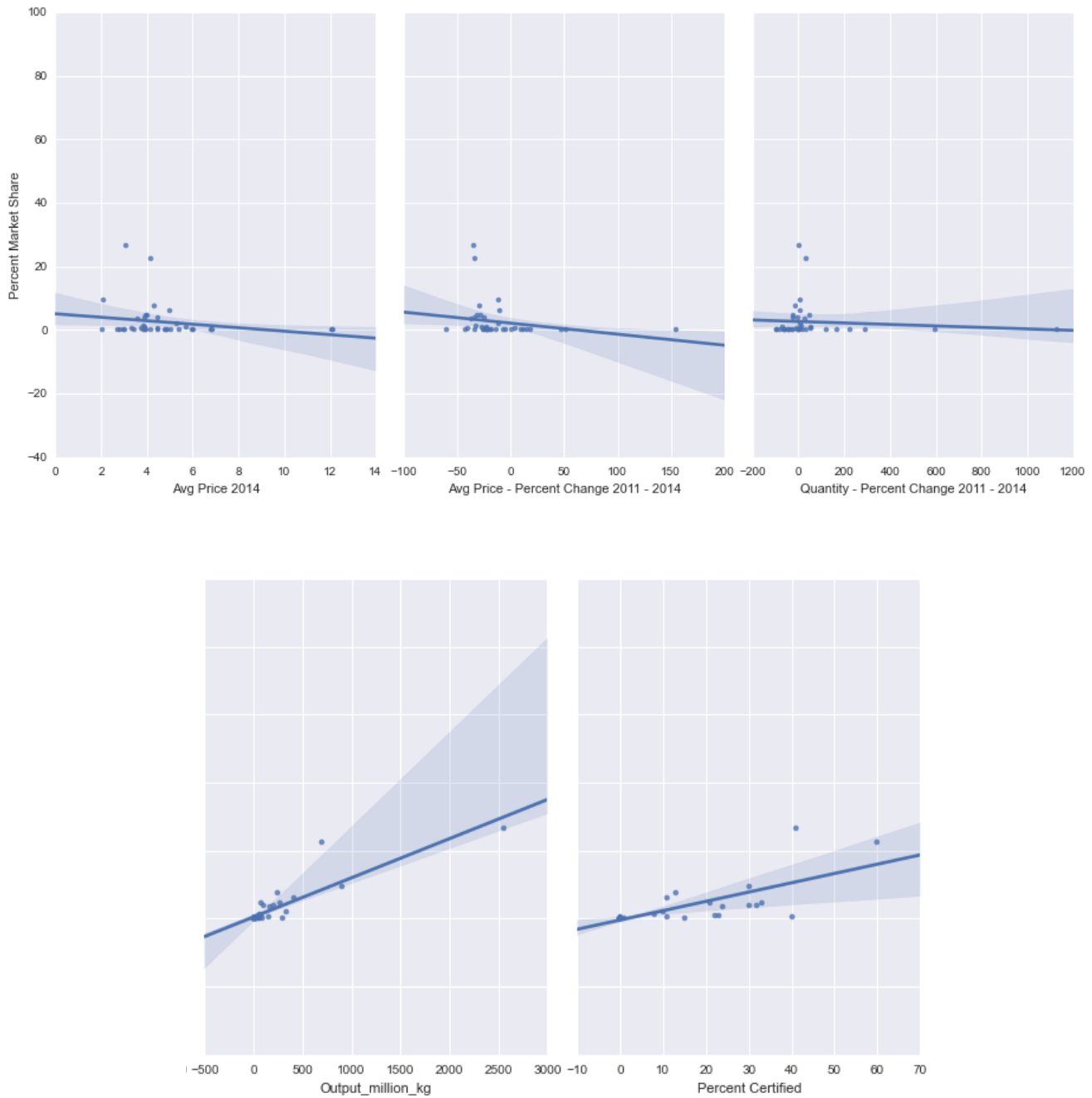
- average price per kg (2014) – custom value divided by
- percent change in average price per kg (2011-2014)
- percent change in quantity of exports (2011-2014)
- total production in kg (2014)
- percent of certified coffee (2012)

The response is market share of individual countries, which is assessed based on the custom value of imported coffee, not the quantity in kg.

As a result, there are 39 countries included in the model:

Brazil, Colombia, Vietnam, Guatemala, Indonesia, Nicaragua, Mexico, Peru, Costa Rica, Honduras, Ethiopia, Papua New Guinea, El Salvador, Kenya, Rwanda, Uganda, Tanzania, Ecuador, China, Panama, India, Bolivia, Democratic Republic of the Congo, Burundi, Cameroon, Laos, Jamaica, Malawi, Yemen, Dominican Republic, Djibouti, Zambia, Republic of the Congo, Haiti, Zimbabwe, Guinea, Madagascar, Nepal, Bosnia-Herzegovina

3. DATA VISUALIZATIONS

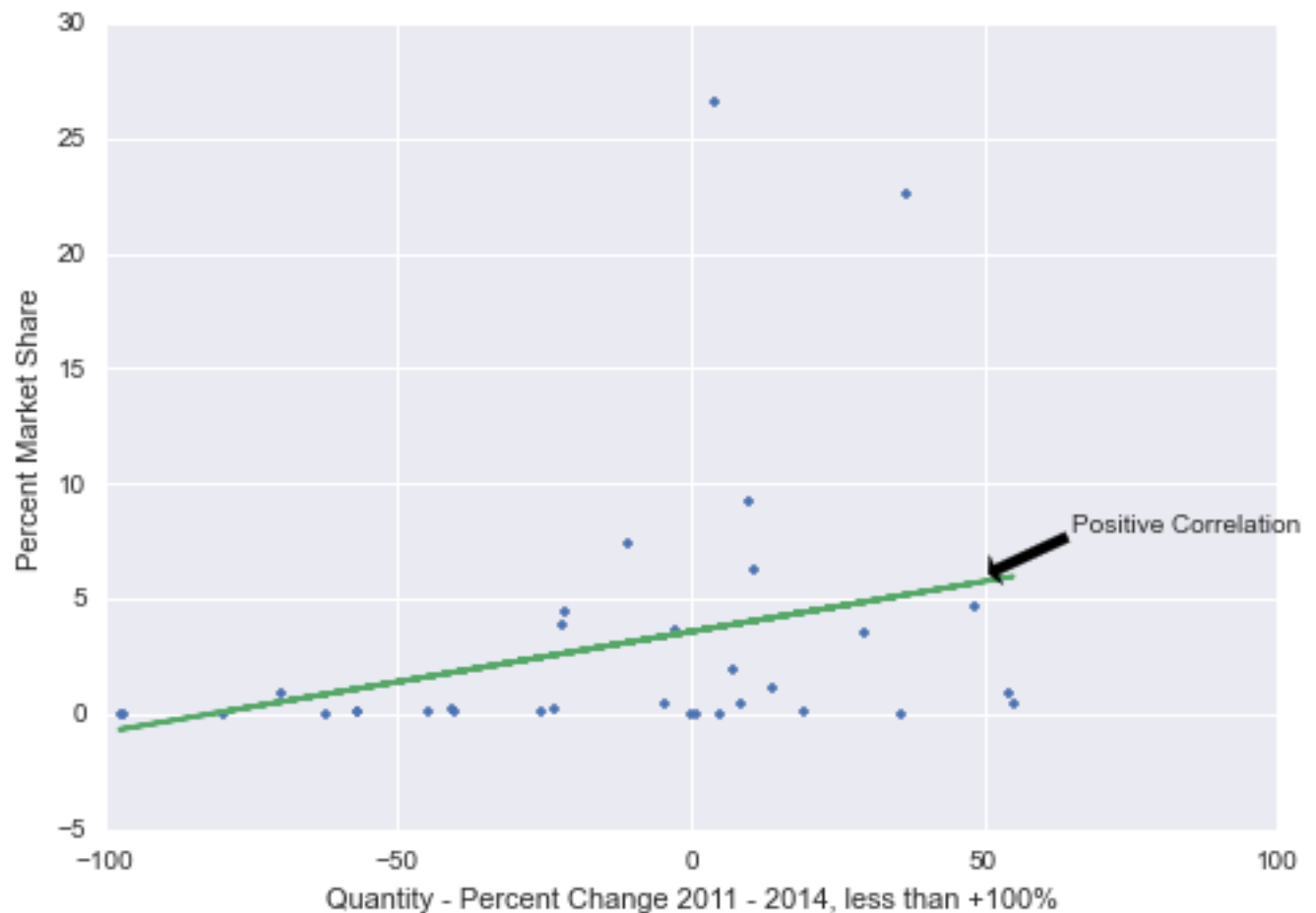


These scatter plots confirm the hypothesis:

- Lower average price, and larger negative change in price 2011-14, results in lower market share.
- Higher national output and higher percentage of certified coffee results in lower market share.
- Except for the change in quantity (2011-14), where growth in supply has a very weak

correlation to market share.

To analyze the percent change in quantity of coffee imported to USA further, data points above +100% have been removed – regarded as outliers; and the remaining 31 data points are plotted below where we can see a positive correlation. However, because most countries in the sample have very low market share and only 10 countries have a market share exceeding 3%, we can only possibly conclude that supply reliability is an important factor for a strong market share. Supply growth, certainly, drives even better market performance.





The correlation matrix above indicates that national output is highly correlated with percent market share. Average price and percent 'certified' follow.

4. MULTIVARIATE LINEAR REGRESSION ANALYSIS

Methods:

1. Train, test, split and cross validation
2. Compare RMSE between various feature combinations (see below)

Trial	FEATURES					RMSE	
	Avg Price 2014	Percent Change - Avg Price 2011-14	Percent Change - Quantity 2011-14	National Output	Percent 'Certified'	Train, Test, Split Method	Cross Validation (mean)
1	✓	✓	✓	✓	✓	2.055	-7.498
2			✓	✓	✓	1.845	-7.493
3				✓	✓	1.69975	-7.540
4				✓		1.523	-7.536
5	✓			✓		1.523	-7.615
6	✓	✓		✓		1.690	-7.487
7	✓		✓	✓		1.662	-7.641
8	✓	✓		✓		1.823	-7.942
9	✓	✓		✓	✓	1.689	-7.506

3. Remove data points/countries with percent change in quantity of coffee imported exceeding +100%, and re-run train, test, split and cross validation

Trial	FEATURES					RMSE	
	Avg Price 2014	Percent Change - Avg Price 2011-14	Percent Change - Quantity 2011-14	National Output	Percent 'Certified'	Train, Test, Split Method	Cross Validation (mean)
1	✓	✓	✓	✓	✓	3.6367	-8.573
2			✓	✓	✓	2.096	-8.054
3			✓	✓		2.086	-7.849

Looking at the first table in this section, the top three models are highlighted in green. The best model (trial #6) is when 'Percent Change – Quantity 2011-14' and 'Percent Certified' are left out. We have discussed the flaws in 'Percent Change – Quantity 2011-14' in Section 3; whereas, 'Percent Certified' data is only found for 18 out of 39 countries (the remaining filled with zeros).

5. CHALLENGES

- Data availability

The availability of national output and percent 'certified' is limited. Only exporters from 20 countries participate in the global certification schemes, and the latest data available is only from 2012. Countries with relatively small production capacity do not make their production data available.

- Dominance of top 10 coffee producing countries

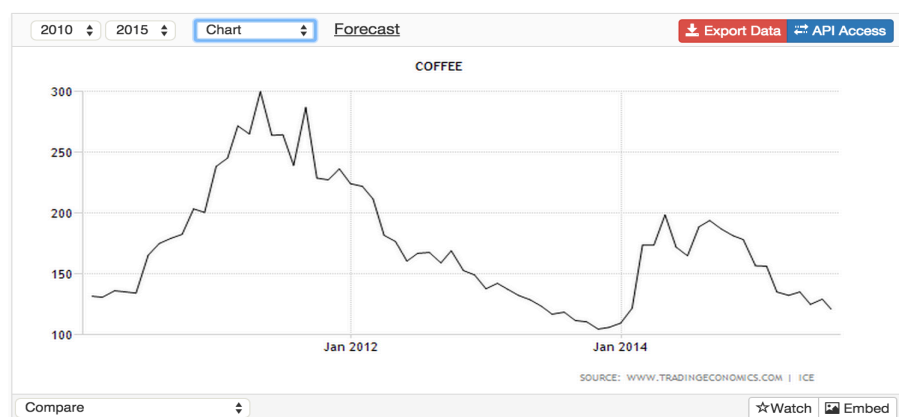
Although 39 countries are included in this study, only 10 countries maintained market share over 3% in 2014. Therefore, the fitting is heavily biased toward these dominant countries.

- Coffee price fluctuations

Prices vary wildly within any one year. By taking an average price at a country level, this study loses some data resolution, not knowing which exporters follow the global coffee prices and which are bound in contracts involving fixed priced.

Coffee 1972-2015 | Data | Chart | Calendar | Forecast | News

Coffee was quoted at 121.03 on Friday July 24. The Coffee averaged 124.57 from 1972 until 2015, reaching an all time high of 339.86 in April of 1977 and a record low of 42.50 in October of 2001.



Source: TradingEconomics.com¹

¹ <http://www.tradingeconomics.com/commodity/coffee>

6. CONCLUSIONS

- Pricing and supply are the top determinants for market share of imported coffee in the US market. By looking at the decline in price between 2011 and 2014, we can see that with greater the magnitude decline, some countries performed better in 2014.
- Sustainability certification is a secondary determinant, with the increasing demand in the US for environmentally and socially 'responsible' as well as 'organic' coffee.
- Because the origin coffee is a major part of coffee branding, a future study can be enhanced with percentage of coffee in the US branded based on its source by country.