

The Market Potential for
Exporting Bottled Wine to
Mainland China (PRC)

The Machine Learning Element

Data Reimagined

SCOPE OF THE ANALYSIS

This analysis was undertaken on behalf of a California company involved both in viticulture and winemaking. With the recent increase in their planted acreage, they wanted to explore the feasibility of expansion into an overseas market, in particular the Chinese one. The owner speaks Chinese and has many connections in mainland China. It was decided to limit the analysis to the potential for bottled exports as opposed to bulk wine shipments since heightened brand visibility was also an objective. This document deals only with the machine learning part of the analysis and does not include all of the other quantitative and qualitative elements that made up the final study.

OBJECTIVES

The machine learning part of the analysis was undertaken to answer the following questions in the context of current Chinese wine consumption. Determining the potential for entering this market was the underlying premise for all queries. As the analysis evolved, other questions were answered but these were the core questions at the outset:

1. Prediction of the global import value of wine by city/province to provide the basis for establishing the US position in the Chinese wine import market.
2. Prediction of the US import value of wine by city/province to provide the basis for determining the best Chinese bottled wine target markets, generating five-year projections of market value, generating five-year projections of market value for a new entrant into the market arena.
3. Prediction of the US import value of wine purchased on the Internet by city/province to provide the basis for determining the best target markets, generating five-year projections of market value, generating five-year projections of market value for a new entrant into the Chinese e-commerce wine sale arena.
4. Profile of Chinese wine drinkers by income and wine preference to derive the most likely target market based on consumer behavior.

The Market Potential for Exporting Bottled Wine to Mainland China – Machine Learning Component

This description of how the analysis was done can be found also together with the full analysis presented to the client, *The Market Potential for Exporting Bottled Wine to Mainland China*. The calculations that were based on this data are part of the full analysis. Data sources are in that document as well.

ASSUMPTIONS AND MODEL INPUTS

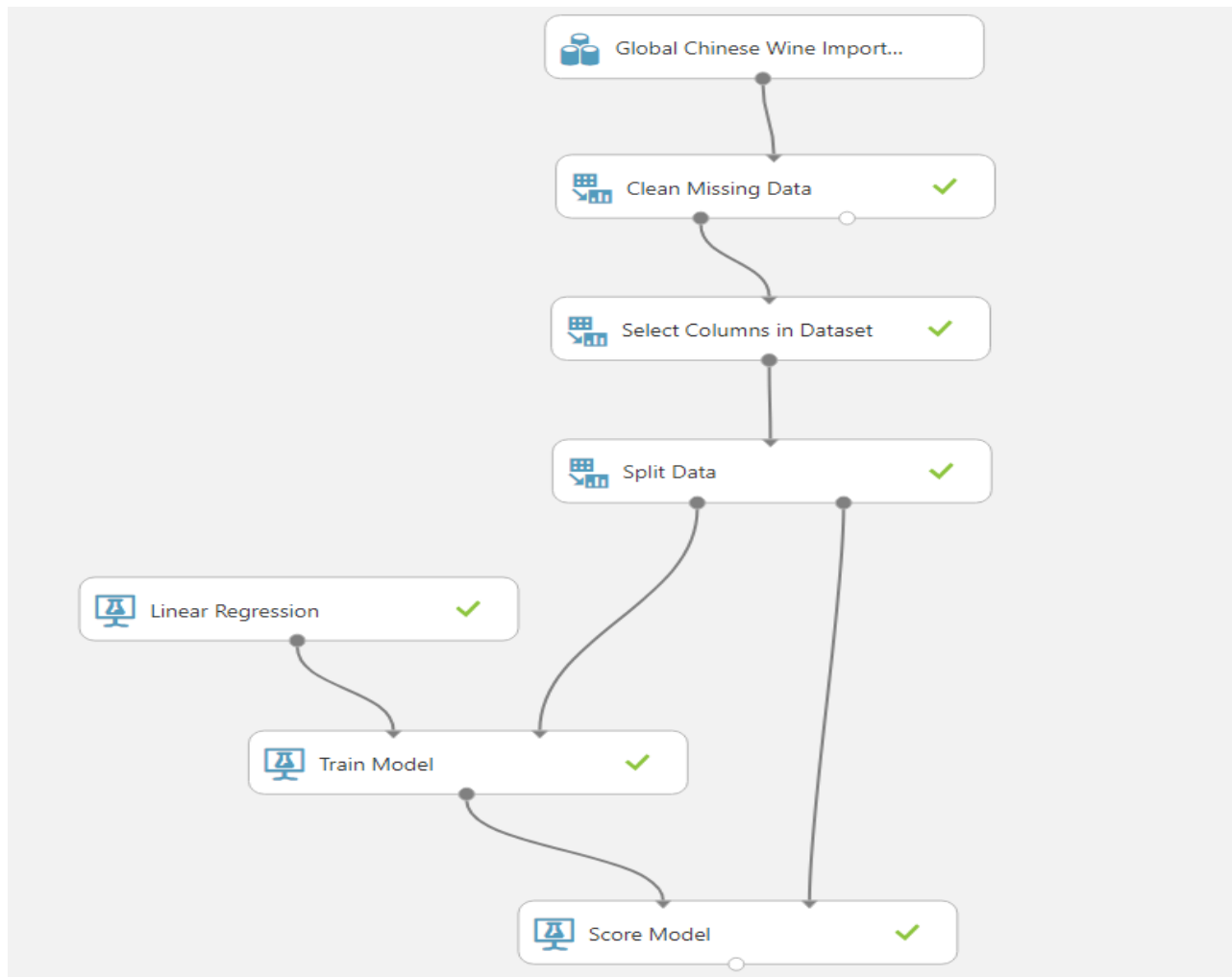
- Import figures were based on Chinese Census Import figures
- Figures for e-commerce bottled wine sales, 20% of the total value, were derived from Amazon China wine sales
- No historical figures existed for individual locations as the basis for revenue or market projections—or projected penetration rates
- Projections regarding individual locations were key to creating a marketing strategy
- Revenue and market projections were critical to making an informed decision about whether or not to enter the market
- Historical data existed for total bottled wine imports but not for city or provincial locations so only 2016/2017 data was used to ensure consistency
- The wine consumer income breakdown was derived from surveys
- Chinese preferences for bottled red, white or sparkling wines—the ones for which figures existed—were available but only for the past year; this reflects the recent exponential increase in bottled wine consumption
- Wine consumed was the same statistically across all age groups between 20 and 60 so this was not considered for inclusion in the model
- To establish global competition for the market, figures for total wine imports by country was calculated over the same period.

METHODOLOGY

In this case, it was decided to construct a model that incorporated a machine learning component with more conventional data analysis methods. The machine learning part of the model was chosen because of the size and nature of the data. The dataset was not large; neither was it comprehensive. Obtaining recent, reliable statistics on Chinese consumption can be difficult.

The Market Potential for Exporting Bottled Wine to Mainland China – Machine Learning Component

While data does exist for total wine imports to China dating back seven years, data did not exist for individual provinces or cities except for 2016/2017. The Chinese market as a whole did have to be validated but it was equally important that the Company's Chinese distributor have a target market framework from which to operate. It seemed reasonable to assume, then, that predictions should be based on 2016/2017 import figures and a number of other wine consumption characteristics to ensure consistency and to try to avoid overfitting or underfitting.



The sampling was small but data did exist for all of China and the individual locations for the following inputs:

- Total Bottled Wine Import Value 2016/2017 by Location
- Total Bottled Wine Import Quantity 2016/2017 by Location

The Market Potential for Exporting Bottled Wine to Mainland China – Machine Learning Component

- Total Chinese Wine Purchased via the Internet 2016/2017 by Location
- Consumer Preference for Bottled Red, White or Sparkling Wine by Location
- Consumption of Bottled Wine by Income Bracket

The same model was used and retrained for each query. The data was split 80-20. A linear regression machine learning algorithm was used.

NOTES ON TRAINED MODEL LOCATIONS

The area numbers in the model correspond to the following locations

Total

Area 1 = Guangdong

Area 2 = Shanghai

Area 3 = Zhejiang

Area 4 = Fujian

Area 5 = Beijing

Area 6 = Tianjin

Area 7 = Shandong

Area 8 = Jiangsu

Area 9 + Liaoning

Area 10 = Sichuan

SCORED MODEL FOR THE VALUE OF US BOTTLED WINE IMPORTED BY CHINA 2016/2017

location	quantity-red	quantity-white	quantity-sparkling	annual-income-14097-26430	annual-income-26430-35242	annual-income-more-than-35,242	import-value-purchased-online	import-value	import-quantity	Scored Labels
area-10	164837.59	13417.01	13417.01	32004.33028	37618.28327	47694.60914	70224.80844	638407.35	115158.01	638407.34907
total	20311927.2	1653296.4	1653296.4	3943697.808	4635470.887	5877114.874	8653373.41	78667031	14190217	78667030.858177
area-6	1021829.03	83172.13	83172.13	198395.0035	233195.9261	295659.1206	435323.9369	3957490.34	713865.08	3957490.329623
area-9	286210.54	23296.21	23296.21	55569.70708	65317.31687	82813.02674	121932.6276	1108478.43	199950.97	1108478.425749
area-5	1116139.09	90848.53	90848.53	216705.9405	254718.8266	322947.0837	475502.3099	4322748.27	779751.51	4322748.273154
area-1	7454595.38	606769.39	606769.39	1447360.024	1701244.766	2156935.328	3175838.352	28871257.74	5207892.14	28871257.699796
area-2	6031743.56	490955.87	490955.87	1171103.737	1376529.73	1745243.05	2569668.985	23360627.14	4213866.52	23360627.085636
area-4	1117779.26	90982.03	90982.03	217024.3913	255093.1376	323421.6566	476201.0642	4329100.58	780897.36	4329100.573019
area-3	1549965.38	126159.97	126159.97	300936.2429	353724.1596	448471.7024	660322.8256	6002934.78	1082829.06	6002934.777984

The Market Potential for Exporting Bottled Wine to Mainland China – Machine Learning Component

SCORED MODEL FOR THE US BOTTLED WINE QUANTITY IMPORTED BY CHINA 2016/2017

rows 9
columns 11

	location	quantity-red	quantity-white	quantity-sparkling	annual-income-14097-26430	annual-income-26430-35242	annual-income-more-than-35,242	import-value-purchased-online	import-value	import-quantity	Scored Labels
view as											
	area-10	164837.59	13417.01	13417.01	32004.33028	37618.28327	47694.60914	70224.80844	638407.35	115158.01	-706106.125
	total	20311927.2	1653296.4	1653296.4	3943697.808	4635470.887	5877114.874	8653373.41	78667031	14190217	35579944
	area-6	1021829.03	83172.13	83172.13	198395.0035	233195.9261	295659.1206	435323.9369	3957490.34	713865.08	837384
	area-9	286210.54	23296.21	23296.21	55569.70708	65317.31687	82813.02674	121932.6276	1108478.43	199950.97	-487506.5
	area-5	1116139.09	90848.53	90848.53	216705.9405	254718.8266	322947.0837	475502.3099	4322748.27	779751.51	1007241.75
	area-1	7454595.38	606769.39	606769.39	1447360.024	1701244.766	2156935.328	3175838.352	28871257.74	5207892.14	12423159
	area-2	6031743.56	490955.87	490955.87	1171103.737	1376529.73	1745243.05	2569668.985	23360627.14	4213866.52	9860522
	area-4	1117779.26	90982.03	90982.03	217024.3913	255093.1376	323421.6566	476201.0642	4329100.58	780897.36	1010195.8125
	area-3	1549965.38	126159.97	126159.97	300936.2429	353724.1596	448471.7024	660322.8256	6002934.78	1082829.06	1788587.375

SCORED MODEL FOR THE VALUE OF CHINESE WINE INTERNET PURCHASES 2016/2017

rows 9
columns 11

	location	quantity-red	quantity-white	quantity-sparkling	annual-income-14097-26430	annual-income-26430-35242	annual-income-more-than-35,242	import-value-purchased-online	import-value	import-quantity	Scored Labels
view as											
	area-10	164837.59	13417.01	13417.01	32004.33028	37618.28327	47694.60914	70224.80844	638407.35	115158.01	-430592.5
	total	20311927.2	1653296.4	1653296.4	3943697.808	4635470.887	5877114.874	8653373.41	78667031	14190217	21697102
	area-6	1021829.03	83172.13	83172.13	198395.0035	233195.9261	295659.1206	435323.9369	3957490.34	713865.08	510647.375
	area-9	286210.54	23296.21	23296.21	55569.70708	65317.31687	82813.02674	121932.6276	1108478.43	199950.97	-297287.625
	area-5	1116139.09	90848.53	90848.53	216705.9405	254718.8266	322947.0837	475502.3099	4322748.27	779751.51	614228.8125
	area-1	7454595.38	606769.39	606769.39	1447360.024	1701244.766	2156935.328	3175838.352	28871257.74	5207892.14	7575801
	area-2	6031743.56	490955.87	490955.87	1171103.737	1376529.73	1745243.05	2569668.985	23360627.14	4213866.52	6013072.5
	area-4	1117779.26	90982.03	90982.03	217024.3913	255093.1376	323421.6566	476201.0642	4329100.58	780897.36	616030.25
	area-3	1549965.38	126159.97	126159.97	300936.2429	353724.1596	448471.7024	660322.8256	6002934.78	1082829.06	1090703.375

The Market Potential for Exporting Bottled Wine to Mainland China – Machine Learning Component

SCORED MODEL FOR THE QUANTITY OF IMPORTED US BOTTLED WINE CONSUMED BY CHINESE WINE DRINKERS WITH INCOMES \$14, 907 TO \$26, 430, 2016/2017

rows 9
columns 11

	location	quantity-red	quantity-white	quantity-sparkling	annual-income-14097-26430	annual-income-26430-35242	annual-income-more-than-35,242	import-value-purchased-online	import-value	import-quantity	Scored Labels
view as											
area-10	164837.59	13417.01	13417.01	32004.33028	37618.28327	47694.60914	70224.80844	638407.35	115158.01	-196238.640625	
total	20311927.2	1653296.4	1653296.4	3943697.808	4635470.887	5877114.874	8653373.41	78667031	14190217	9888260	
area-6	1021829.03	83172.13	83172.13	198395.0035	233195.9261	295659.1206	435323.9369	3957490.34	713865.08	232723.015625	
area-9	286210.54	23296.21	23296.21	55569.70708	65317.31687	82813.02674	121932.6276	1108478.43	199950.97	-135486.1875	
area-5	1116139.09	90848.53	90848.53	216705.9405	254718.8266	322947.0837	475502.3099	4322748.27	779751.51	279929.28125	
area-1	7454595.38	606769.39	606769.39	1447360.024	1701244.766	2156935.328	3175838.352	28871257.74	5207892.14	3452603.25	
area-2	6031743.56	490955.87	490955.87	1171103.737	1376529.73	1745243.05	2569668.985	23360627.14	4213866.52	2740404.25	
area-4	1117779.26	90982.03	90982.03	217024.3913	255093.1376	323421.6566	476201.0642	4329100.58	780897.36	280750.25	
area-3	1549965.38	126159.97	126159.97	300936.2429	353724.1596	448471.7024	660322.8256	6002934.78	1082829.06	497078.3125	

SCORED MODEL FOR THE QUANTITY OF IMPORTED US BOTTLED WINE CONSUMED BY CHINESE WINE DRINKERS WITH INCOMES \$26, 430 TO \$35, 242, 2016/2017

rows 9
columns 11

	location	quantity-red	quantity-white	quantity-sparkling	annual-income-14097-26430	annual-income-26430-35242	annual-income-more-than-35,242	import-value-purchased-online	import-value	import-quantity	Scored Labels
view as											
area-10	164837.59	13417.01	13417.01	32004.33028	37618.28327	47694.60914	70224.80844	638407.35	115158.01	-230661.3125	
total	20311927.2	1653296.4	1653296.4	3943697.808	4635470.887	5877114.874	8653373.41	78667031	14190217	11622779	
area-6	1021829.03	83172.13	83172.13	198395.0035	233195.9261	295659.1206	435323.9369	3957490.34	713865.08	273545.40625	
area-9	286210.54	23296.21	23296.21	55569.70708	65317.31687	82813.02674	121932.6276	1108478.43	199950.97	-159252.0625	
area-5	1116139.09	90848.53	90848.53	216705.9405	254718.8266	322947.0837	475502.3099	4322748.27	779751.51	329032.28125	
area-1	7454595.38	606769.39	606769.39	1447360.024	1701244.766	2156935.328	3175838.352	28871257.74	5207892.14	4058231.5	
area-2	6031743.56	490955.87	490955.87	1171103.737	1376529.73	1745243.05	2569668.985	23360627.14	4213866.52	3221104	
area-4	1117779.26	90982.03	90982.03	217024.3913	255093.1376	323421.6566	476201.0642	4329100.58	780897.36	329997.28125	
area-3	1549965.38	126159.97	126159.97	300936.2429	353724.1596	448471.7024	660322.8256	6002934.78	1082829.06	584271.875	

The Market Potential for Exporting Bottled Wine to Mainland China – Machine Learning Component

SCORED MODEL FOR THE QUANTITY OF IMPORTED US BOTTLED WINE CONSUMED BY CHINESE WINE DRINKERS WITH INCOMES OVER \$35, 242, 2016/2017

rows	columns											
9	11	location	quantity-red	quantity-white	quantity-sparkling	annual-income-14097-26430	annual-income-26430-35242	annual-income-more-than-35,242	import-value-purchased-online	import-value	import-quantity	Scored Labels
view as												
		area-10	164837.59	13417.01	13417.01	32004.33028	37618.28327	47694.60914	70224.80844	638407.35	115158.01	-292445.625
		total	20311927.2	1653296.4	1653296.4	3943697.808	4635470.887	5877114.874	8653373.41	78667031	14190217	14736027
		area-6	1021829.03	83172.13	83172.13	198395.0035	233195.9261	295659.1206	435323.9369	3957490.34	713865.08	346816.59375
		area-9	286210.54	23296.21	23296.21	55569.70708	65317.31687	82813.02674	121932.6276	1108478.43	199950.97	-201908.96875
		area-5	1116139.09	90848.53	90848.53	216705.9405	254718.8266	322947.0837	475502.3099	4322748.27	779751.51	417166
		area-1	7454595.38	606769.39	606769.39	1447360.024	1701244.766	2156935.328	3175838.352	28871257.74	5207892.14	5145259.5
		area-2	6031743.56	490955.87	490955.87	1171103.737	1376529.73	1745243.05	2569668.985	23360627.14	4213866.52	4083900.75
		area-4	1117779.26	90982.03	90982.03	217024.3913	255093.1376	323421.6566	476201.0642	4329100.58	780897.36	418389.5
		area-3	1549965.38	126159.97	126159.97	300936.2429	353724.1596	448471.7024	660322.8256	6002934.78	1082829.06	740773.375

From this model, projections were made that enabled in part a determination of the viability of this company entering the Chinese bottled wine market.

HOW THE IMPORT VALUE AND QUANTITY INFORMATION WAS USED

This part of the model was used for two purposes: To eliminate locations that were not deemed to be potential target markets on the basis of import value and quantity; and, to identify the locations that appeared to be potential target markets on the basis of import value and quantity. Two of the original ten locations became the test part of the model.

The results effectively eliminated two locations, Sichuan and Liaoning, as potential target markets because their scored numbers were consistently negative in all categories. They were not used for any projections made on the basis of this model. This left six locations that were used as the potential target market.

The scored figures of locations with positive numbers, which showed reasonable results, were used as the basis for projections.

*The Market Potential for Exporting Bottled Wine to Mainland China –
Machine Learning Component*

	US Wine Import Value		US Wine Import Quantity		Value of Chinese Internet Wine Purchases	
	Input	Scored	Input	Scored	Input	Scored
Total	\$78,667,031	\$197,246,336	14,190,217	35,579,944	\$8,653,373	\$21,697,102
Sichuan	\$638,407	-\$3,914,476	115,158	-706,106	\$70,225	-\$430,593
Tianjin	\$3,957,490	\$4,642,249	713,865	837,384	\$435,324	\$510,647
Liaoning	\$118,478	-\$2,702,615	199,951	-487,507	\$121,933	-\$297,288
Beijing	\$4,322,748	\$5,583,898	779,752	1,007,241	\$475,502	\$614,229
Guangdong	\$28,871,258	\$68,870,904	5,207,892	1,243,159	\$3,175,838	\$7,575,801
Shanghai	\$23,360,627	\$54,664,284	4,213,867	9,860,522	\$2,569,669	\$6,013,073
Fujian	\$4,329,101	\$5,600,275	780,897	1,010,195	\$476,201	\$616,030
Zhejiang	\$6,002,935	\$9,915,483	1,082,829	1,788,587	\$660,323	\$1,090,703

HOW THE IMPORT VALUE AND QUANTITY INFORMATION WAS USED

This part of the model was used for two purposes: To eliminate locations that were not deemed to be potential target markets on the basis of consumer income; and, to identify the locations that did appear to be potential target markets on the basis of consumer income.

Consistent with the value and quantity figures, scored results eliminated Sichuan and Liaoning as potential target markets because their numbers were again negative in all categories. They were not used for any projections made on the basis of this model. This left six locations that were used as the potential target market. The scored figures of locations with positive numbers were used for projections.

	Chinese Wine Drinkers with Incomes \$14907 to \$26,430		Chinese Wine Drinkers with Incomes \$26,430 to \$35,242		Chinese Wine Drinkers with Incomes Over \$35,242	
	Input	Scored	Input	Scored	Input	Scored
Total	3,943,698	9,888,260	4,635,471	11,622,779	5,877,115	14,736,027
Sichuan	32,004	-3,914,476	37,618	-230,661	47,695	-292,445
Tianjin	198,395	232,273	233,196	273,545	295,659	346,817
Liaoning	55,570	-135,486	65,317	-159,252	82,813	-201,909
Beijing	216,706	279,929	254,719	329,032	322,947	417,166
Guangdong	1,447,360	3,452,603	1,701,245	4,058,231	2,156,935	5,145,259
Shanghai	1,171,104	2,740,404	1,376,530	3,221,104	1,745,243	4,083,900
Fujian	217,024	280,750	255,093	329,997	323,422	418,389
Zhejiang	300,936	497,078	353,724	584,271	448,472	740,773