Final Case Study

Julia Cuellar

DSC 550

Final CS

Description:

A synthetic dataset that describes pizza sales for a pizza place somewhere in the US. While the contents are artificial, the ingredients used to make the pizzas are far from it. There are 32 different pizzas that fall into 4 different categories: *classic* (classic pizzas: 'You probably had one like it before, but never like this!'), *chicken* (pizzas with chicken as a major ingredient: 'Try the Southwest Chicken Pizza! You'll love it!'), *supreme* (pizzas that try a little harder: 'My Soppressata pizza uses only the finest salami from my personal salumist!'), and *veggie* (pizzas without any meats whatsoever: 'My Five Cheese pizza has so many cheeses, I can only offer it in Large Size!').

Usage:

pizzaplace

Format:

A tibble with 49574 rows and 7 variables:

id

The ID for the order, which consists of one or more pizzas at a given **date** and **time**

date

A character representation of the **order** date, expressed in the ISO 8601 date format (YYYY-MM-DD)

time

A character representation of the **order** time, expressed as a 24-hour time the ISO 8601 extended time format (hh:mm:ss)

name

The short name for the pizza

size

The size of the pizza, which can either be **S**, **M**, **L**, **XL** (rare!), or **XXL** (even rarer!); most pizzas are available in the **S**, **M**, and **L** sizes but exceptions apply

type

The category or type of pizza, which can either be **classic**, **chicken**, **supreme**, or **veggie**

price

The price of the pizza and the amount that it sold for (in USD)

Details:

Each pizza in the dataset is identified by a short name. The following listings provide the full names of each pizza and their main ingredients.

Classic Pizzas:

- *classic_dlx*: The Classic Deluxe Pizza (Pepperoni, Mushrooms, Red Onions, Red Peppers, Bacon)
- *big_meat*: The Big Meat Pizza (Bacon, Pepperoni, Italian Sausage, Chorizo Sausage)
- *pepperoni*: The Pepperoni Pizza (Mozzarella Cheese, Pepperoni)
- *hawaiian*: The Hawaiian Pizza (Sliced Ham, Pineapple, Mozzarella Cheese)
- *pep_msh_pep*: The Pepperoni, Mushroom, and Peppers Pizza (Pepperoni, Mushrooms, and Green Peppers)
- *ital_cpcllo*: The Italian Capocollo Pizza (Capocollo, Red Peppers, Tomatoes, Goat Cheese, Garlic, Oregano)
- *napolitana*: The Napolitana Pizza (Tomatoes, Anchovies, Green Olives, Red Onions, Garlic)
- *the_greek*: The Greek Pizza (Kalamata Olives, Feta Cheese, Tomatoes, Garlic, Beef Chuck Roast, Red Onions)

Chicken Pizzas:

- *thai_ckn*: The Thai Chicken Pizza (Chicken, Pineapple, Tomatoes, Red Peppers, Thai Sweet Chilli Sauce)
- *bbq_ckn*: The Barbecue Chicken Pizza (Barbecued Chicken, Red Peppers, Green Peppers, Tomatoes, Red Onions, Barbecue Sauce)
- *southw_ckn*: The Southwest Chicken Pizza (Chicken, Tomatoes, Red Peppers, Red Onions, Jalapeno Peppers, Corn, Cilantro, Chipotle Sauce)
- *cali_ckn*: The California Chicken Pizza (Chicken, Artichoke, Spinach, Garlic, Jalapeno Peppers, Fontina Cheese, Gouda Cheese)
- *ckn_pesto*: The Chicken Pesto Pizza (Chicken, Tomatoes, Red Peppers, Spinach, Garlic, Pesto Sauce)
- *ckn_alfredo*: The Chicken Alfredo Pizza (Chicken, Red Onions, Red Peppers, Mushrooms, Asiago Cheese, Alfredo Sauce)

Supreme Pizzas:

- *brie_carre*: The Brie Carre Pizza (Brie Carre Cheese, Prosciutto, Caramelized Onions, Pears, Thyme, Garlic)
- *calabrese*: The Calabrese Pizza (‘Nduja Salami, Pancetta, Tomatoes, Red Onions, Friggitello Peppers, Garlic)
- *soppressata*: The Soppressata Pizza (Soppressata Salami, Fontina Cheese, Mozzarella Cheese, Mushrooms, Garlic)
- *sicilian*: The Sicilian Pizza (Coarse Sicilian Salami, Tomatoes, Green Olives, Luganega Sausage, Onions, Garlic)
- *ital_supr*: The Italian Supreme Pizza (Calabrese Salami, Capocollo, Tomatoes, Red Onions, Green Olives, Garlic)
- *peppr_salami:* The Pepper Salami Pizza (Genoa Salami, Capocollo, Pepperoni, Tomatoes, Asiago Cheese, Garlic)
- *prsc_argla*: The Prosciutto and Arugula Pizza (Prosciutto di San Daniele, Arugula, Mozzarella Cheese)
- *spinach_supr*: The Spinach Supreme Pizza (Spinach, Red Onions, Pepperoni, Tomatoes, Artichokes, Kalamata Olives, Garlic, Asiago Cheese)
- *spicy_ital*: The Spicy Italian Pizza (Capocollo, Tomatoes, Goat Cheese, Artichokes, Peperoncini verdi, Garlic)

Vegetable Pizzas

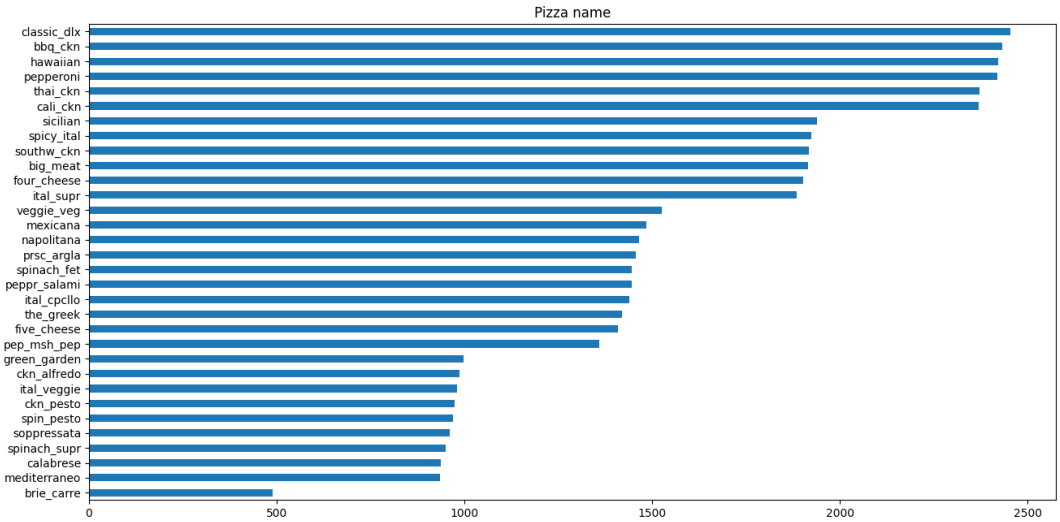
- *mexicana*: The Mexicana Pizza (Tomatoes, Red Peppers, Jalapeno Peppers, Red Onions, Cilantro, Corn, Chipotle Sauce, Garlic)
- *four_cheese*: The Four Cheese Pizza (Ricotta Cheese, Gorgonzola Piccante Cheese, Mozzarella Cheese, Parmigiano Reggiano Cheese, Garlic)
- *five_cheese*: The Five Cheese Pizza (Mozzarella Cheese, Provolone Cheese, Smoked Gouda Cheese, Romano Cheese, Blue Cheese, Garlic)
- *spin_pesto*: The Spinach Pesto Pizza (Spinach, Artichokes, Tomatoes, Sun-dried Tomatoes, Garlic, Pesto Sauce)
- *veggie_veg*: The Vegetables + Vegetables Pizza (Mushrooms, Tomatoes, Red Peppers, Green Peppers, Red Onions, Zucchini, Spinach, Garlic)
- *green_garden*: The Green Garden Pizza (Spinach, Mushrooms, Tomatoes, Green Olives, Feta Cheese)
- *mediterraneo*: The Mediterranean Pizza (Spinach, Artichokes, Kalamata Olives, Sun-dried Tomatoes, Feta Cheese, Plum Tomatoes, Red Onions)
- *spinach_fet*: The Spinach and Feta Pizza (Spinach, Mushrooms, Red Onions, Feta Cheese, Garlic)
- *ital_veggie*: The Italian Vegetables Pizza (Eggplant, Artichokes, Tomatoes, Zucchini, Red Peppers, Garlic, Pesto Sauce)

<u>Original Pizza Place Data</u>

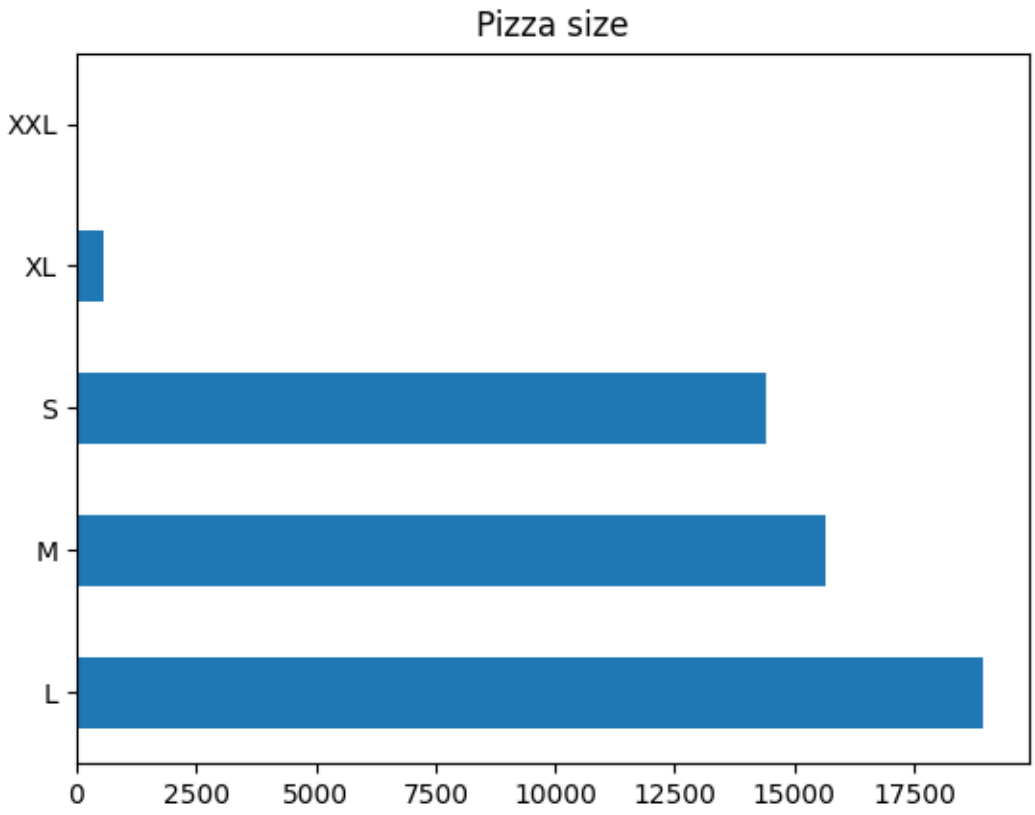
	Unnamed: 0	id	date	...	size	type	price
0	1	2015-000001	2015-01-01	...	M	classic	13.25

1	2	2015-000002	2015-01-01	...	M	classic	16.00
2	3	2015-000002	2015-01-01	...	M	veggie	16.00
3	4	2015-000002	2015-01-01	...	L	chicken	20.75
4	5	2015-000002	2015-01-01	...	L	veggie	18.50

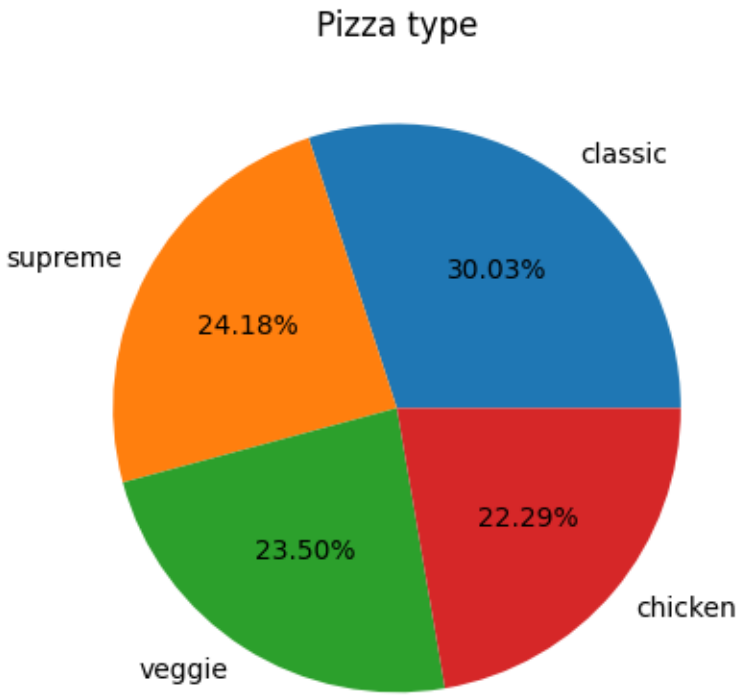
Pizza Name Bar Chart



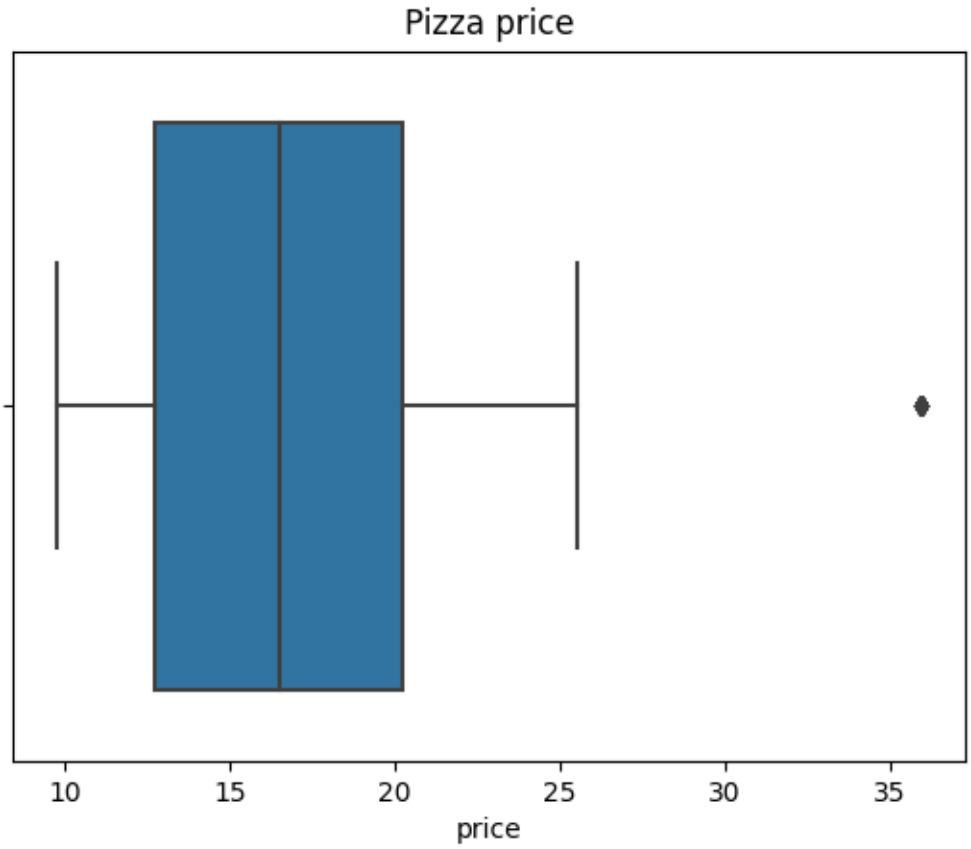
Pizza Size Bar Chart



Pizza Type Pie Chart



Pizza Price Boxplot



Updated Format:

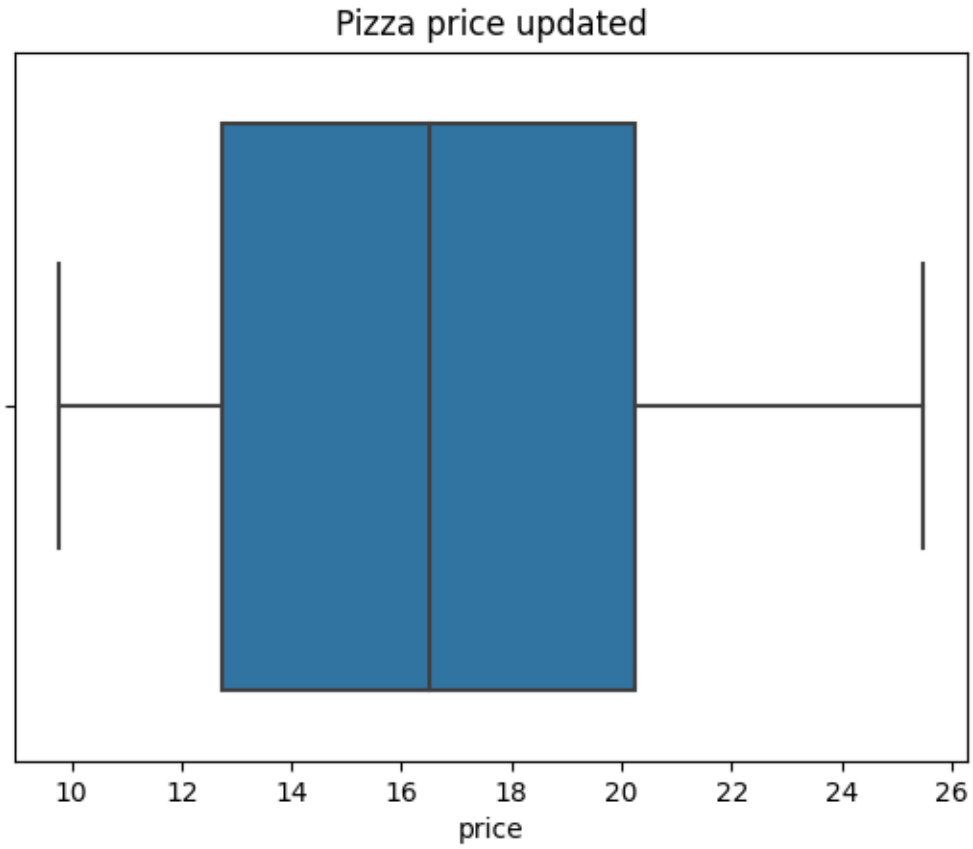
A tibble with 49546 rows and 5 variables

Updated Pizza Place Data

	time	name	size	type	price
0	11:38:36	hawaiian	M	classic	13.25
1	11:57:40	classic_dlx	M	classic	16.00
2	11:57:40	mexicana	M	veggie	16.00
3	11:57:40	thai_ckn	L	chicken	20.75
4	11:57:40	five_cheese	L	veggie	18.50

EDA:

I dropped the unnamed, **id**, and **date** columns because they were irrelevant from the year of 2015. I also removed the rows of prices that ranged from $35 or greater because they were outliers.



Statistical questions:

1) Is there a relationship between the size of a pizza versus the type of pizza purchased?
2) Is there a relationship between the name of a pizza versus the price of pizza purchased?

Hypotheses:

1. There will be more large classic pizzas purchased.
2. There will be more classic dlx pizzas purchased at $12 or $16.

Multiple linear regression model for size of pizza vs type of pizza purchased:

OLS Regression Results

==========
==========

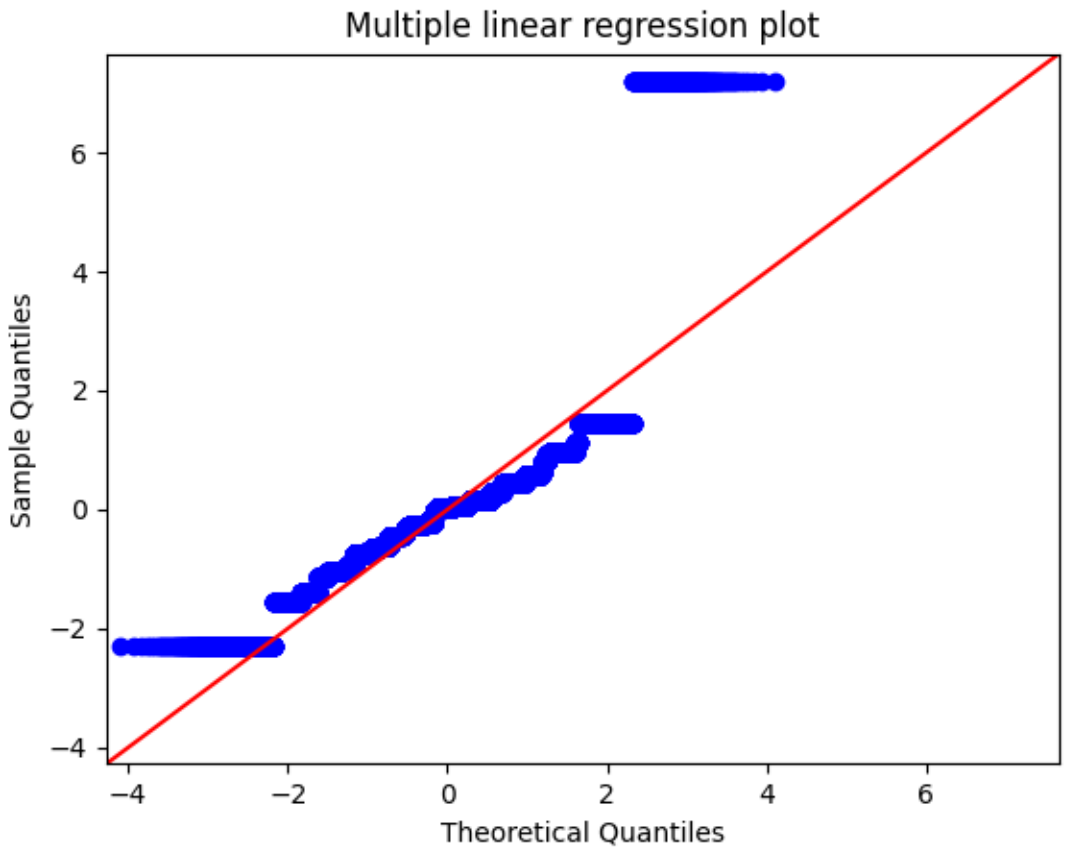
Dep. Variable:	price	R-squared:	0.847
Model:	OLS	Adj. R-squared:	0.847
Method:	Least Squares	F-statistic:	4.570e+04
Date:	Wed, 17 Feb 2021	Prob (F-statistic):	0.00
Time:	14:44:08	Log-Likelihood:	-87171.
No. Observations:	49546	AIC:	1.744e+05
Df Residuals:	49539	BIC:	1.744e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	20.5052	0.015	1343.493	0.000	20.475	20.535
C(size)[T.M]	-3.8127	0.015	-250.543	0.000	-3.843	-3.783
C(size)[T.S]	-7.1114	0.016	-448.156	0.000	-7.143	-7.080
C(size)[T.XL]	7.0274	0.062	113.857	0.000	6.906	7.148
C(type)[T.classic]	-2.0326	0.018	-112.266	0.000	-2.068	-1.997
C(type)[T.supreme]	0.1538	0.019	8.277	0.000	0.117	0.190
C(type)[T.veggie]	-1.0795	0.019	-57.800	0.000	-1.116	-1.043

Omnibus:	38426.463	Durbin-Watson:	2.006
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1319807.442
Skew:	3.452	Prob(JB):	0.00
Kurtosis:	27.324	Cond. No.	11.7

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.



Cross table of size of pizza vs type of pizza purchased:

type	chicken	classic	supreme	veggie
size				
L	4932	4057	4564	5403
M	3894	4112	4046	3583
S	2224	6139	3377	2663
XL	0	552	0	0

Simple linear regression model for name of pizza vs price of pizza purchased:

OLS Regression Results

==========

Dep. Variable: price R-squared: 0.315

Model: OLS Adj. R-squared: 0.314

Method: Least Squares F-statistic: 733.8

Date: Wed, 17 Feb 2021 Prob (F-statistic): 0.00

Time: 14:44:17 Log-Likelihood: -1.2431e+05

No. Observations: 49546 AIC: 2.487e+05

Df Residuals: 49514 BIC: 2.490e+05

Df Model: 31

Covariance Type: nonrobust

==========

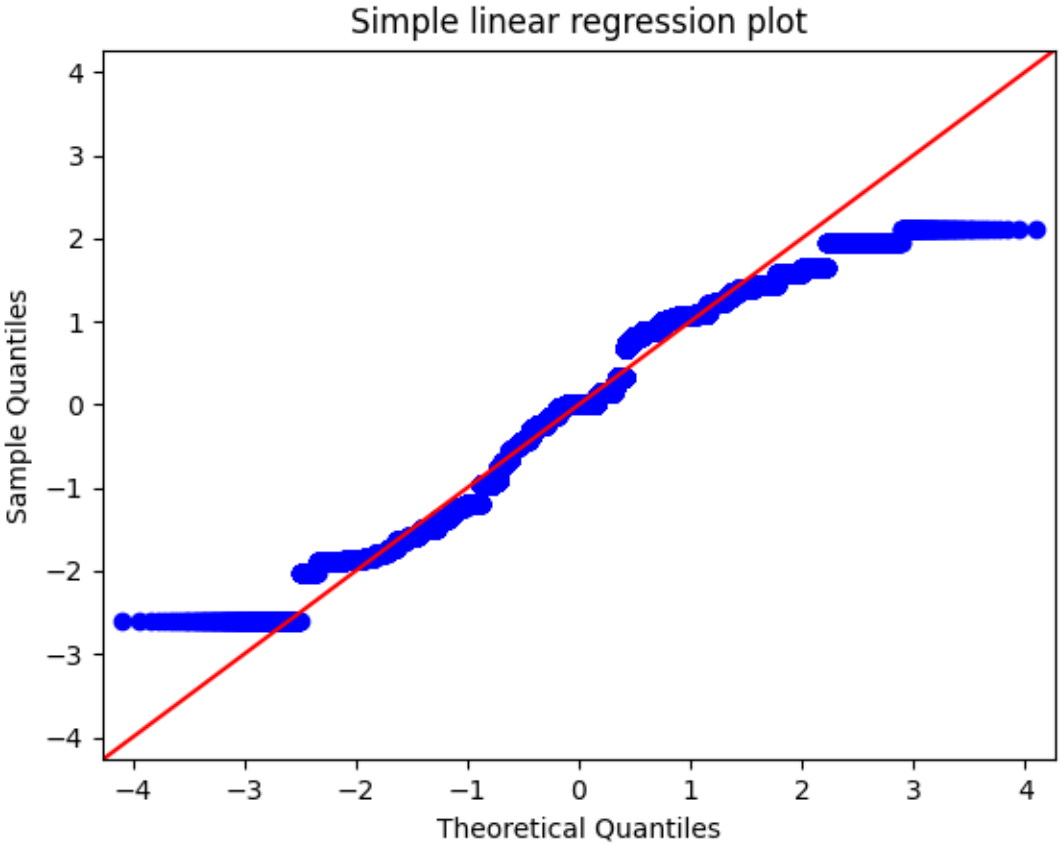
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	17.5855	0.060	291.468	0.000	17.467	17.704
C(name)[T.big_meat]	-5.5855	0.091	-61.436	0.000	-5.764	-5.407
C(name)[T.brie_carre]	6.0645	0.147	41.161	0.000	5.776	6.353
C(name)[T.calabrese]	-0.5799	0.114	-5.069	0.000	-0.804	-0.356
C(name)[T.cali_ckn]	-0.1132	0.086	-1.318	0.188	-0.281	0.055
C(name)[T.ckn_alfredo]	-0.4627	0.112	-4.120	0.000	-0.683	-0.243
C(name)[T.ckn_pesto]	-0.4203	0.113	-3.724	0.000	-0.642	-0.199
C(name)[T.classic_dlx]	-2.0207	0.085	-23.733	0.000	-2.188	-1.854
C(name)[T.five_cheese]	0.9145	0.100	9.180	0.000	0.719	1.110
C(name)[T.four_cheese]	-0.6214	0.091	-6.823	0.000	-0.800	-0.443
C(name)[T.green_garden]	-3.5878	0.112	-32.065	0.000	-3.807	-3.368
C(name)[T.hawaiian]	-4.2605	0.085	-49.881	0.000	-4.428	-4.093
C(name)[T.ital_cpcllo]	-0.1349	0.099	-1.363	0.173	-0.329	0.059
C(name)[T.ital_supr]	0.1834	0.091	2.009	0.045	0.004	0.362

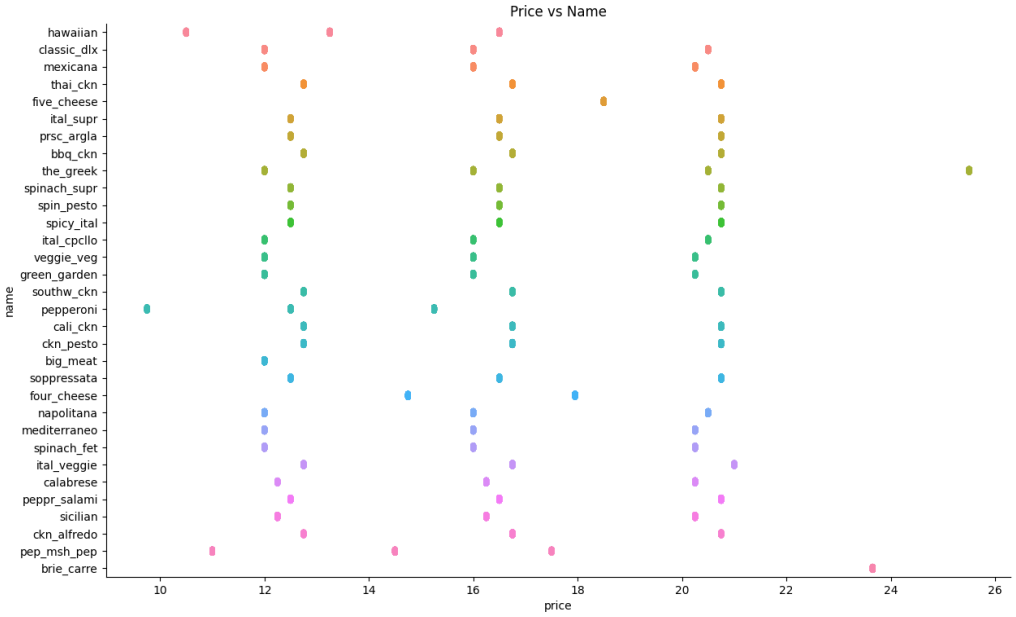
C(name)[T.ital_veggie]	-1.2560	0.113	-11.161	0.000	-1.477	-1.035
C(name)[T.mediterraneo]	-1.1396	0.115	-9.950	0.000	-1.364	-0.915
C(name)[T.mexicana]	0.4608	0.098	4.702	0.000	0.269	0.653
C(name)[T.napolitana]	-1.1327	0.098	-11.508	0.000	-1.326	-0.940
C(name)[T.pep_msh_pep]	-3.7264	0.101	-36.980	0.000	-3.924	-3.529
C(name)[T.pepperoni]	-5.1117	0.085	-59.821	0.000	-5.279	-4.944
C(name)[T.peppr_salami]	0.0694	0.099	0.702	0.483	-0.124	0.263
C(name)[T.prsc_argla]	-0.9807	0.099	-9.949	0.000	-1.174	-0.787
C(name)[T.sicilian]	-1.6204	0.091	-17.885	0.000	-1.798	-1.443
C(name)[T.soppressata]	-0.4932	0.113	-4.350	0.000	-0.715	-0.271
C(name)[T.southw_ckn]	0.5187	0.091	5.707	0.000	0.341	0.697
C(name)[T.spicy_ital]	0.5180	0.091	5.706	0.000	0.340	0.696
C(name)[T.spin_pesto]	-1.5072	0.113	-13.339	0.000	-1.729	-1.286
C(name)[T.spinach_fet]	-1.4920	0.099	-15.100	0.000	-1.686	-1.298
C(name)[T.spinach_supr]	-1.5037	0.114	-13.209	0.000	-1.727	-1.281
C(name)[T.thai_ckn]	0.7334	0.086	8.541	0.000	0.565	0.902
C(name)[T.the_greek]	2.1325	0.100	21.325	0.000	1.937	2.329
C(name)[T.veggie_veg]	-1.6126	0.097	-16.596	0.000	-1.803	-1.422

Omnibus:	4627.087	Durbin-Watson:	1.993
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1698.323
Skew:	-0.211	Prob(JB):	0.00
Kurtosis:	2.197	Cond. No.	26.7

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.


Simple linear regression plot
Sample Quantiles
Theoretical Quantiles
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3
4


Price vs Name
name
price
hawaiian
classic_dlx
mexicana
thai_ckn
five_cheese
ital_supr
prsc_argla
bbq_ckn
the_greek
spinach_supr
spin_pesto
spicy_ital
ital_cpcllo
veggie_veg
green_garden
southw_ckn
pepperoni
cali_ckn
ckn_pesto
big_meat
soppressata
four_cheese
napolitana
mediterraneo
spinach_fet
ital_veggie
calabrese
peppr_salami
sicilian
ckn_alfredo
pep_msh_pep
brie_carre
10
12
14
16
18
20
22
24
26

Cross table of price of pizza vs name of pizza purchased:

name	bbq_ckn	big_meat	brie_carre	...	thai_ckn	the_greek	veggie_veg
price				...			
9.75	0	0	0	...	0	0	0
10.50	0	0	0	...	0	0	0
11.00	0	0	0	...	0	0	0
12.00	0	1914	0	...	0	304	464
12.25	0	0	0	...	0	0	0
12.50	0	0	0	...	0	0	0
12.75	484	0	0	...	480	0	0
13.25	0	0	0	...	0	0	0
14.50	0	0	0	...	0	0	0
14.75	0	0	0	...	0	0	0
15.25	0	0	0	...	0	0	0
16.00	0	0	0	...	0	281	635
16.25	0	0	0	...	0	0	0
16.50	0	0	0	...	0	0	0
16.75	956	0	0	...	481	0	0
17.50	0	0	0	...	0	0	0
17.95	0	0	0	...	0	0	0
18.50	0	0	0	...	0	0	0
20.25	0	0	0	...	0	0	427
20.50	0	0	0	...	0	255	0
20.75	992	0	0	...	1410	0	0
21.00	0	0	0	...	0	0	0
23.65	0	0	490	...	0	0	0
25.50	0	0	0	...	0	552	0

[24 rows x 32 columns]

Conclusion:

A regression model was made for both size of pizza vs type of pizza purchased, and name of pizza vs price of pizza purchased. For the first regression model (multiple linear), the relationship between size and type of pizza is concluded with more small classic pizzas were purchased. For the second regression model (simple linear), the relationship between name and price of pizza is concluded with more classic_dlx pizzas purchased at $12.

Summary:

From the conclusion of the first regression model (small classic pizzas) for size of pizza vs type of pizza purchased, it can be determined that customers would prefer to purchase smaller pizzas with meats and several different vegetables. From the conclusion of the second regression model (classic_dlx pizzas purchased at $12) for name of pizza vs price of pizza purchased, it can be determined that customers would prefer to purchase classic_dlx pizzas at $12.

Final results:

Customers at this pizza place would prefer purchasing small classic deluxe pizzas that are worth $12. This classic pizza encompasses the ingredients of pepperoni, mushrooms, red onions, red peppers, and bacon. If the trend of customers is purchasing the smallest size of pizza with a mix of meats and vegetables at the somewhat cheapest price for a pizza at this pizza place, then the pizza place is doing well off for what they charge as well as the variety of their menu.

R: A year of pizza sales from a pizza place (vincentarelbundock.github.io)