Boost up Your Certification Score

CompTIA
DY0-001
CompTIA DataX Certification Exam

For More Information – Visit link below:

https://www.examsboost.com/

Product Version

✓ Up to Date products, reliable and verified.
✓ Questions and Answers in PDF Format.

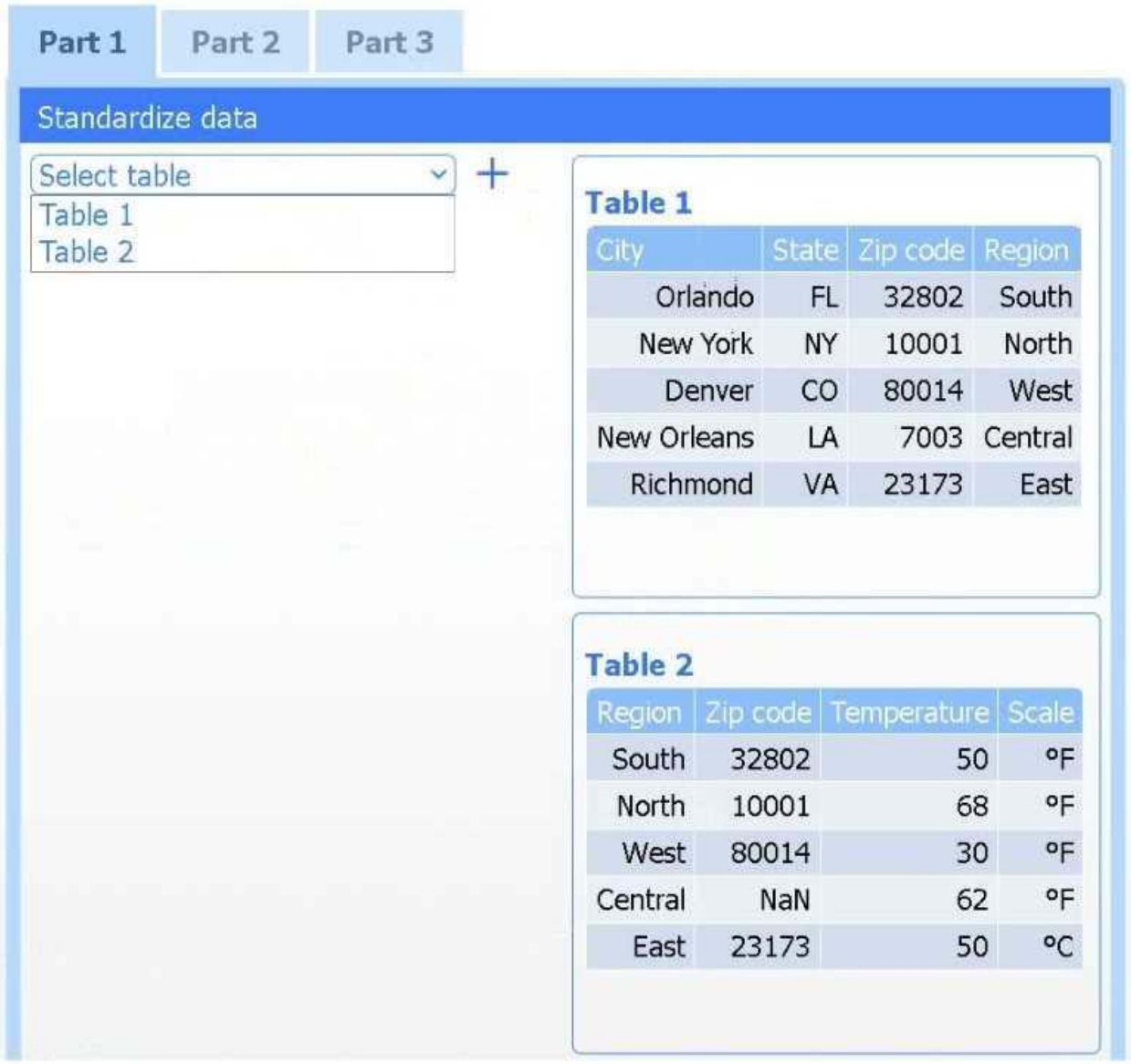
Visit us at: https://www.examsboost.com/test/dy0-001

Latest Version: 6.0

Question: 1

SIMULATION
A client has gathered weather data on which regions have high temperatures. The client would like a visualization to gain a better understanding of the data.
INSTRUCTIONS
Part 1
Review the charts provided and use the drop-down menu to select the most appropriate way to standardize the data.
Part 2
Answer the questions to determine how to create one data set.
Part 3
Select the most appropriate visualization based on the data set that represents what the client is looking for.
If at any time you would like to bring back the initial state of the simulation, please click the Reset All button.

Visit us at: https://www.examsboost.com/test/dy0-001



Visit us at: https://www.examsboost.com/test/dy0-001

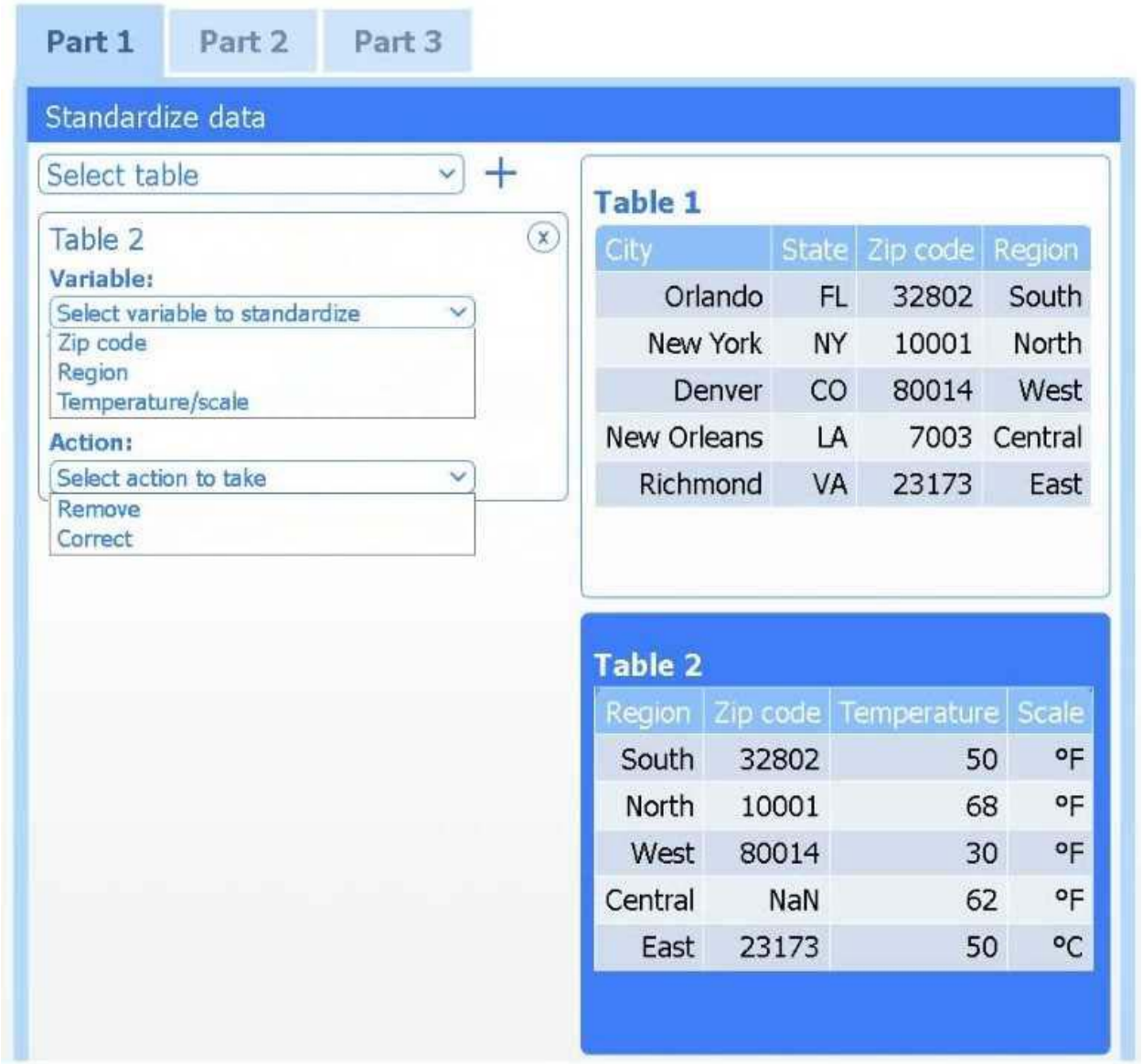
Visit us at: https://www.examsboost.com/test/dy0-001

Visit us at: https://www.examsboost.com/test/dy0-001

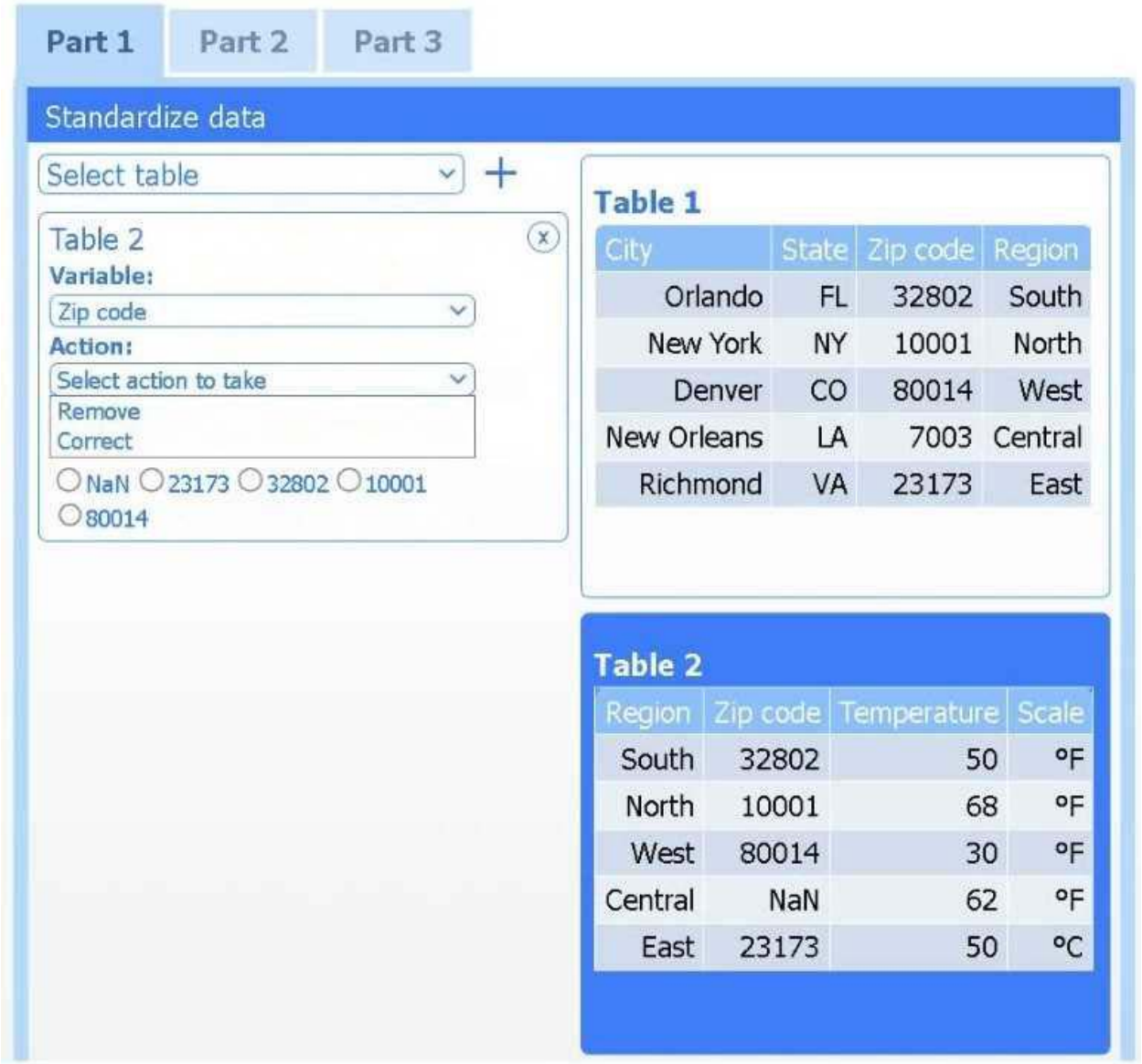
Visit us at: https://www.examsboost.com/test/dy0-001

Visit us at: https://www.examsboost.com/test/dy0-001

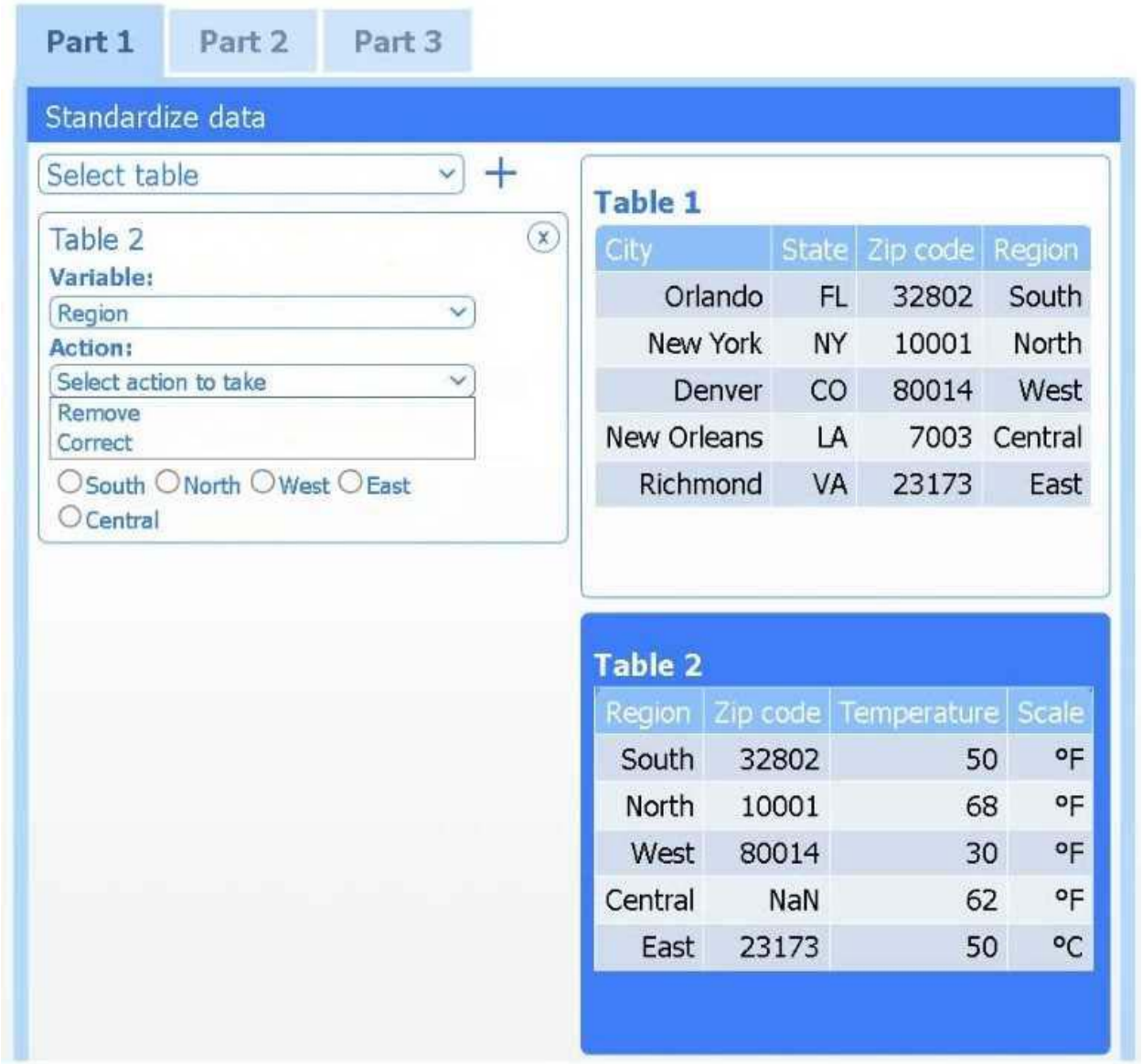
Visit us at: https://www.examsboost.com/test/dy0-001



Visit us at: https://www.examsboost.com/test/dy0-001



Visit us at: https://www.examsboost.com/test/dy0-001



Visit us at: https://www.examsboost.com/test/dy0-001

Visit us at: https://www.examsboost.com/test/dy0-001

Part 1 | Part 2 | Part 3

Merge data

Select the **most** appropriate method to use when combining these two tables:

- ○ Data matching
- ○ Filter
- ○ Union
- ○ Deduplication

Select the **most** appropriate variable to use when joining these sets of data:

- ○ Region
- ○ Zip code

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

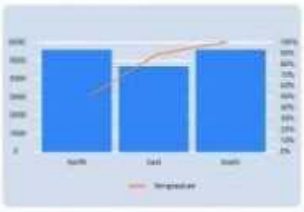
Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

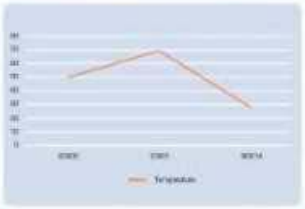
Visit us at: https://www.examsboost.com/test/dy0-001

Part 1 | Part 2 | **Part 3**

Visualization

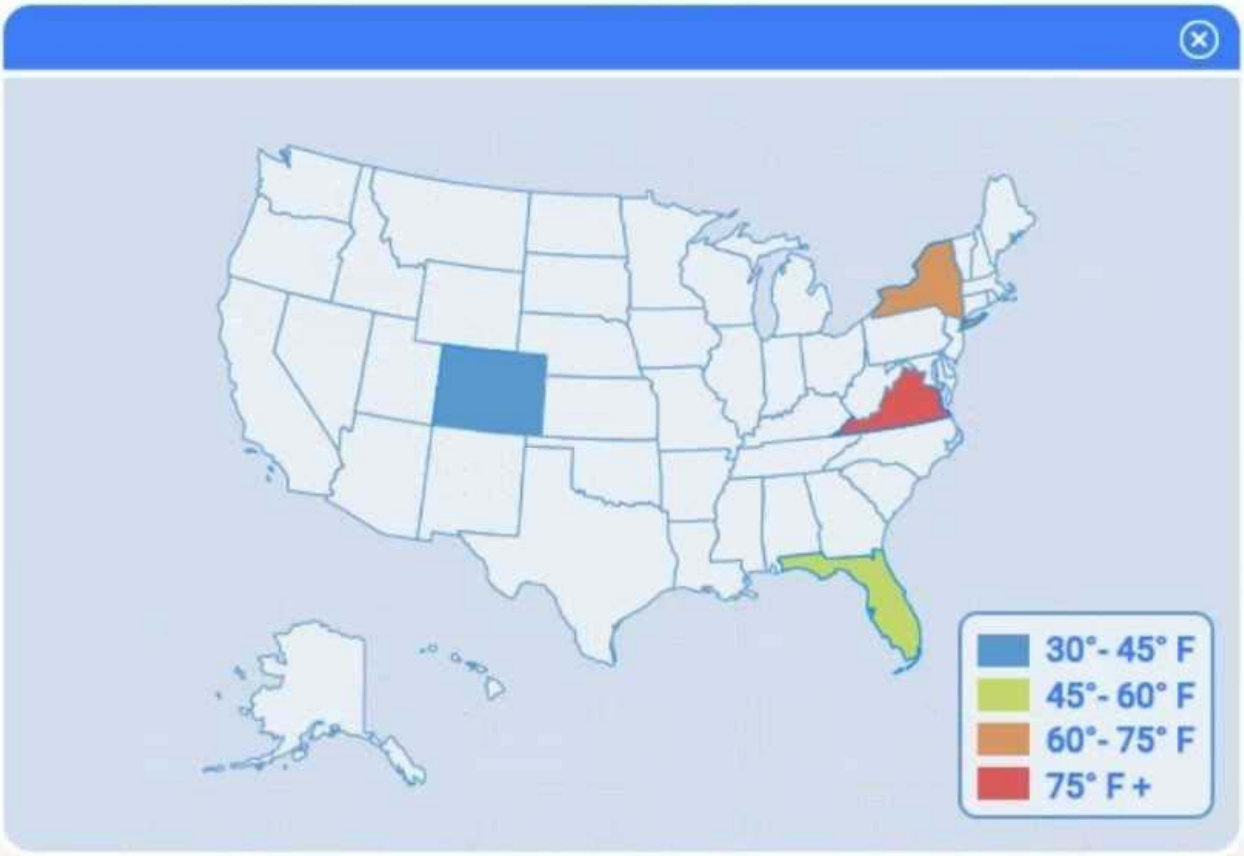
Select the **most** appropriate visualization based on the data set which represents what the client is looking for:

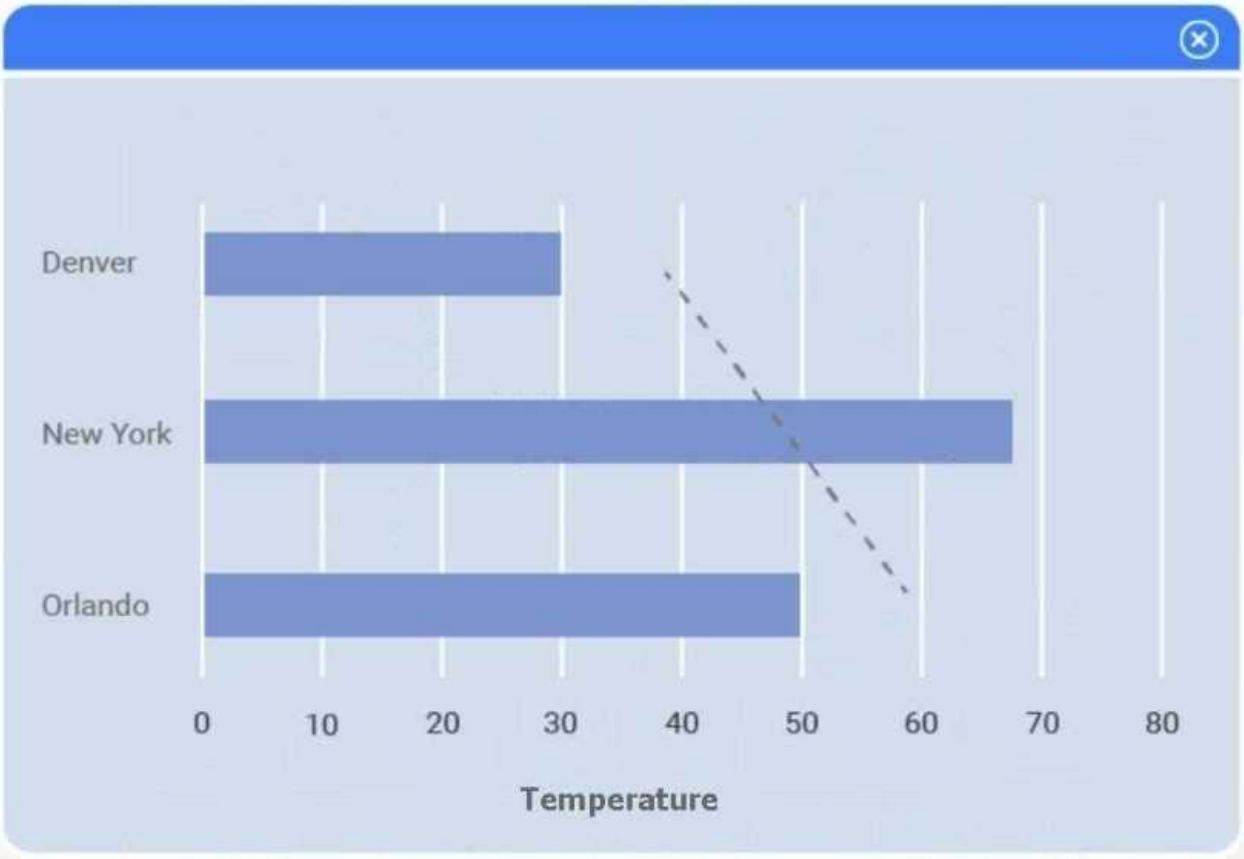




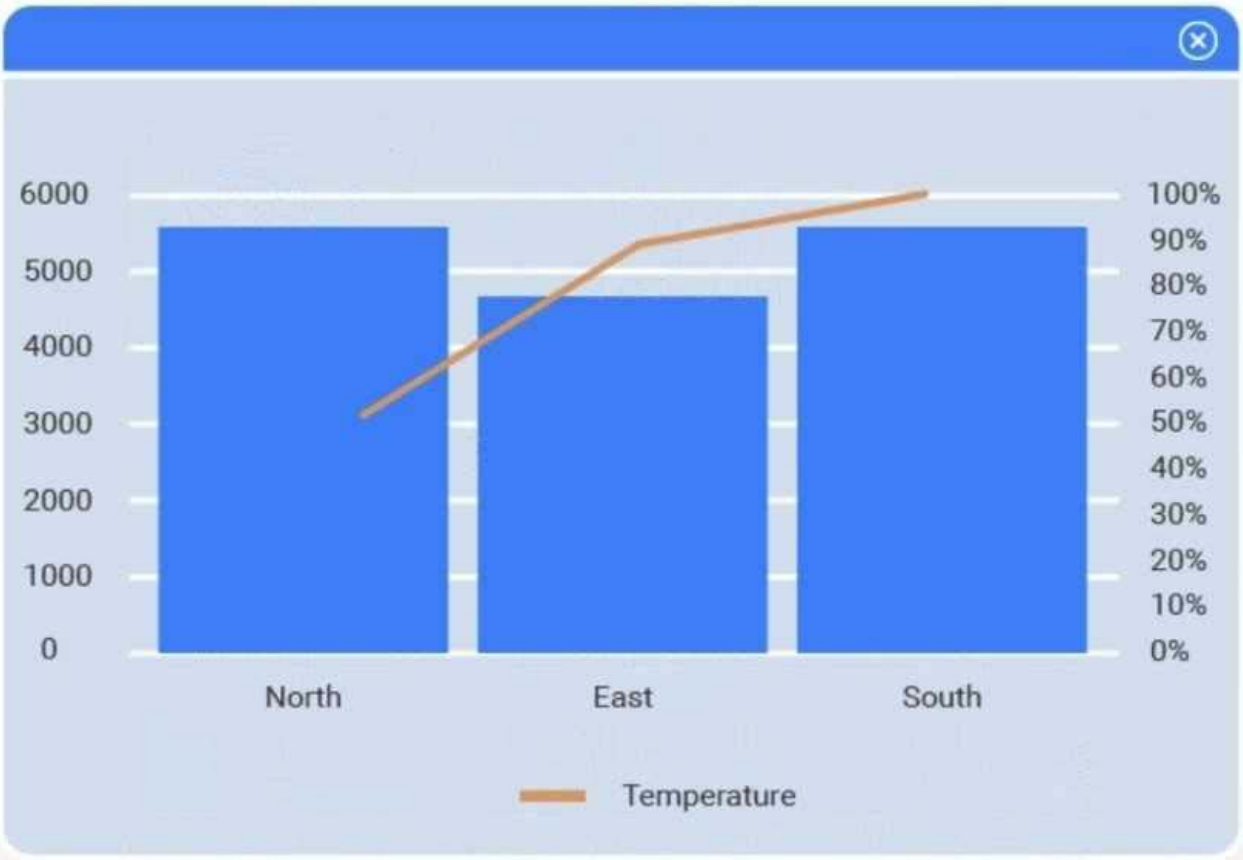
Region	City	State	Zip code	Temperature	Scale
South	Orlando	FL	32802	50	°F
North	New York	NY	10001	68	°F
West	Denver	CO	80014	30	°F
Central	New Orleans	LA	7003		
East	Richmond	VA	23173	50	°C
Central			NaN	62	°F

Visit us at: https://www.examsboost.com/test/dy0-001

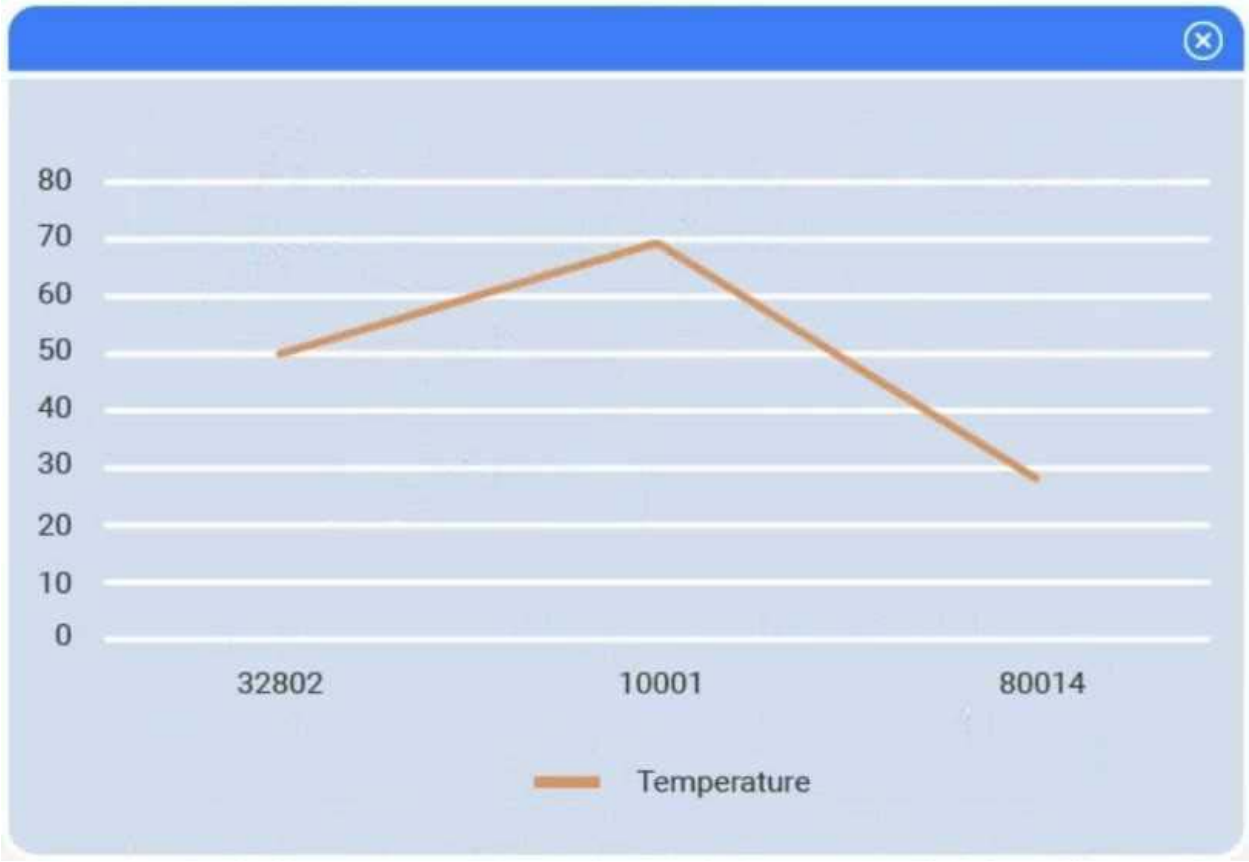




Visit us at: https://www.examsboost.com/test/dy0-001

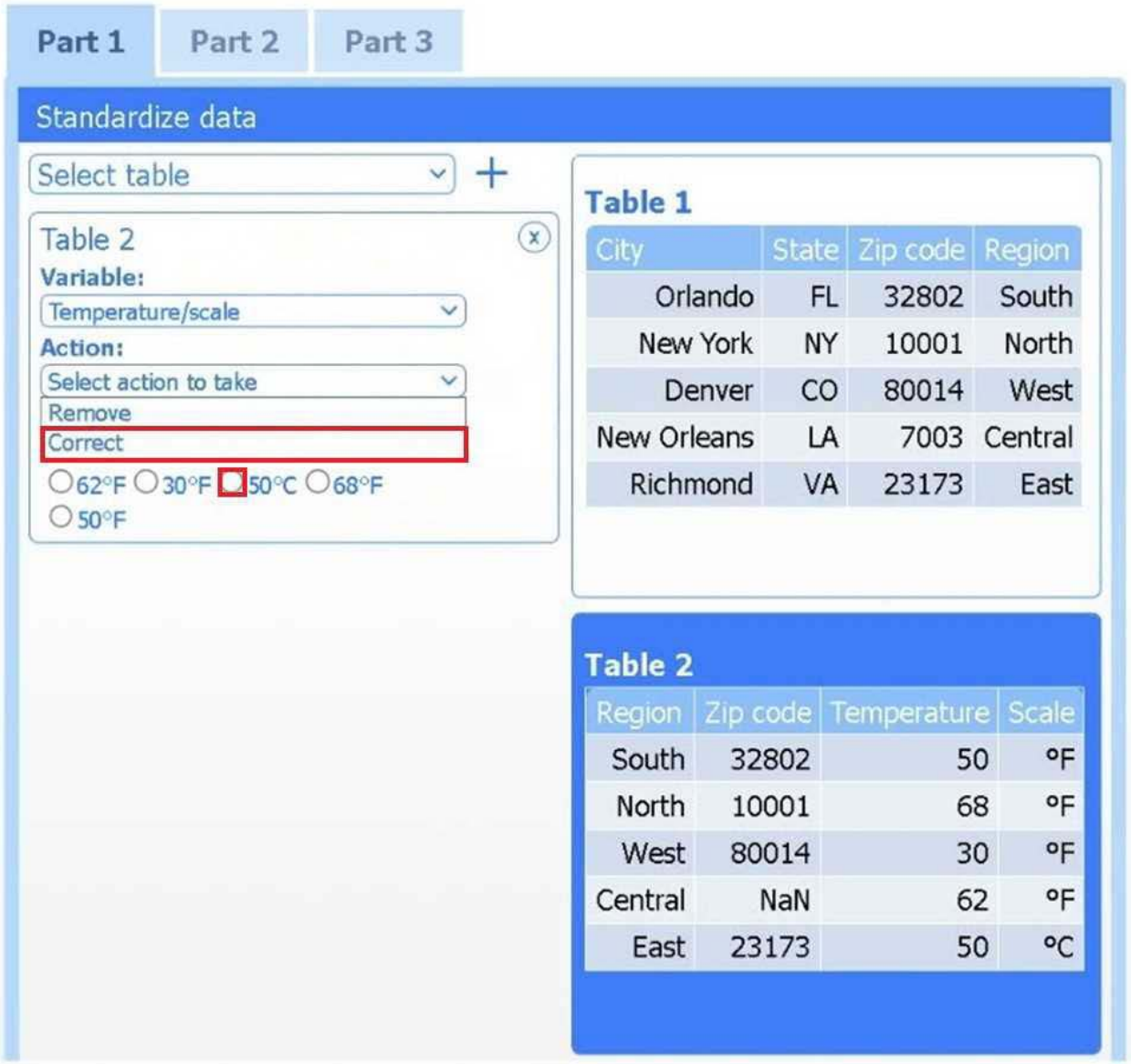

6000
5000
4000
3000
2000
1000
0
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
North
East
South
Temperature

68°F
62°F
122°F
73°F
50°F



Answer: See explanation below.

Explanation:
Explanation:
Part 1
Select Table 2. Table 2 contains mixed temperature scales (°F and °C) that must be standardized before visualization.
Variable: Temperature/scale
Action: Correct
Value to correct: 50 °C



Part 2
Method: Data matching
Join variable: Zip code
You need to merge the two tables by aligning matching records, which is a data-matching (join) operation, and ZIP code is the shared, uniquely identifying field linking each region's weather reading to its city.

Visit us at: https://www.examsboost.com/test/dy0-001

Part 3

Choose the choropleth map (the first option).

A choropleth map best shows geographic variation in temperature by coloring each state (or region) according to its recorded value. This lets the client immediately see where the highest and lowest temperatures occur across the U.S. without distracting elements like bubble size or combined chart axes.

Visit us at: https://www.examsboost.com/test/dy0-001

Part 1 | Part 2 | Part 3

Visualization

Select the **most** appropriate visualization based on the data set which represents what the client is looking for:

Region	City	State	Zip code	Temperature	Scale
South	Orlando	FL	32802	50	°F
North	New York	NY	10001	68	°F
West	Denver	CO	80014	30	°F
Central	New Orleans	LA	7003		
East	Richmond	VA	23173	50	°C
Central			NaN	62	°F

Question: 2

SIMULATION

A data scientist needs to determine whether product sales are impacted by other contributing factors. The client has provided the data scientist with sales and other variables in the data set. The data scientist decides to test potential models that include other information.

INSTRUCTIONS

Part 1

Use the information provided in the table to select the appropriate regression model.

Part 2

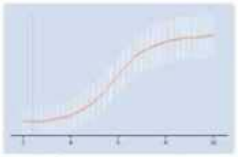
Review the summary output and variable table to determine which variable is statistically significant.

If at any time you would like to bring back the initial state of the simulation, please click the Reset All button.

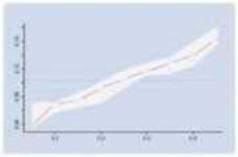
Visit us at: https://www.examsboost.com/test/dy0-001

Part 1 | Part 2

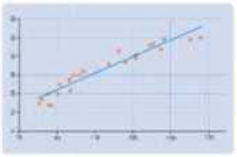
Given the R^2 values, which of the following regression models **best** fits the relationship between the variables?



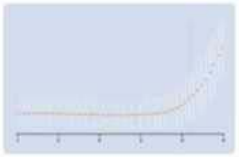
Ridge regression
R^2 0.5



Quantile regression
R^2 0.6



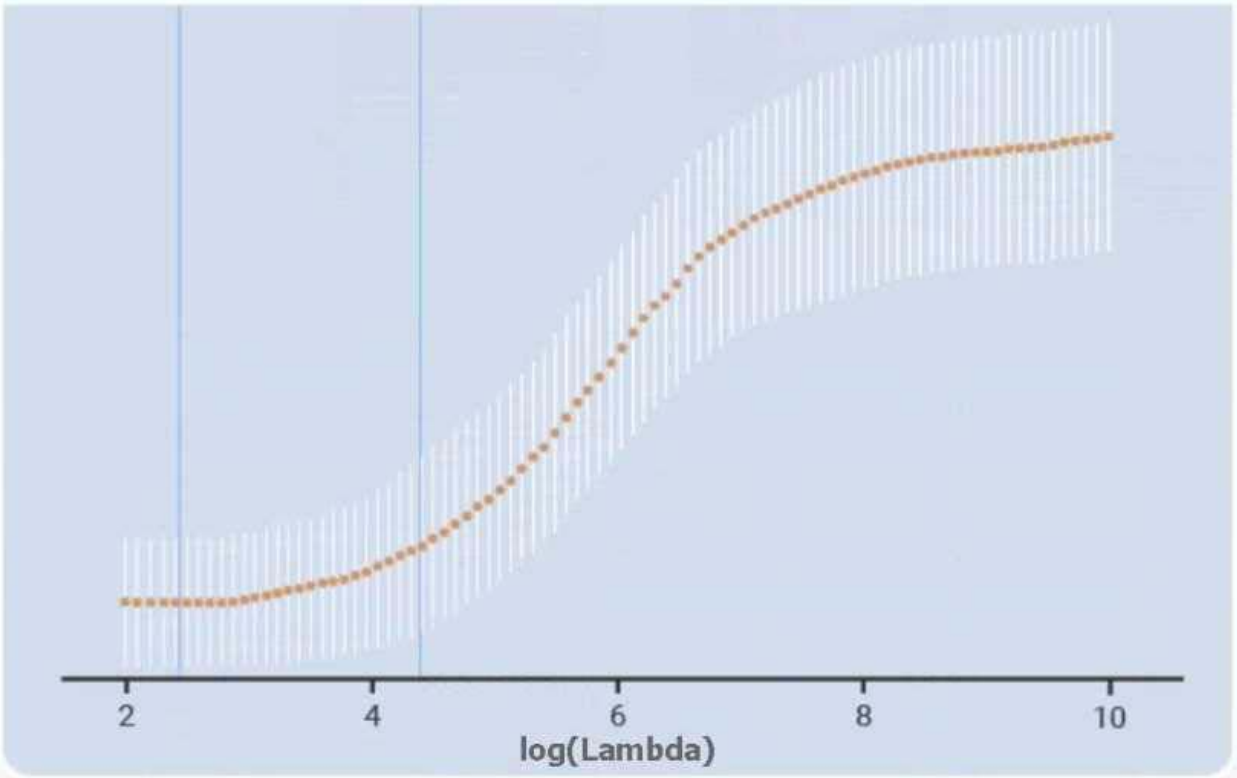
Linear regression
R^2 0.8



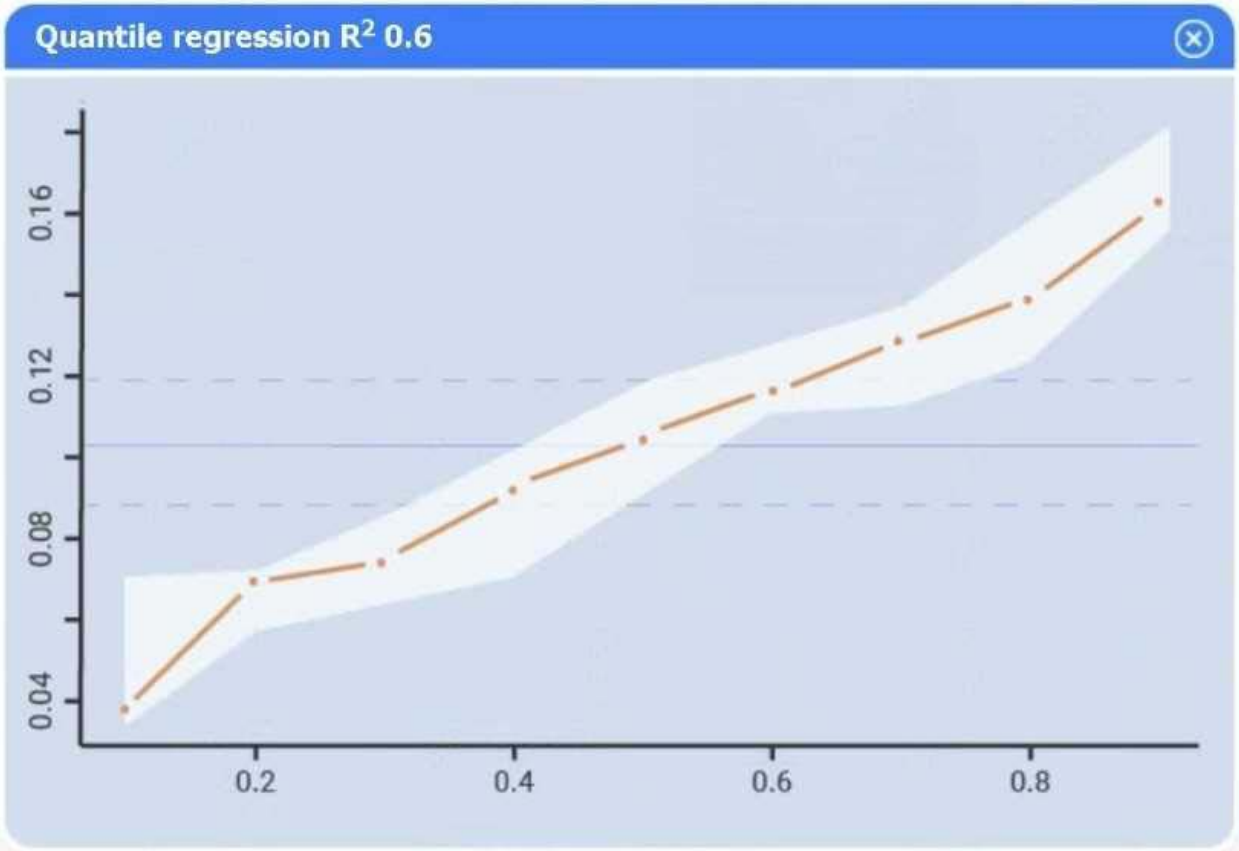
Lasso regression
R^2 0.62

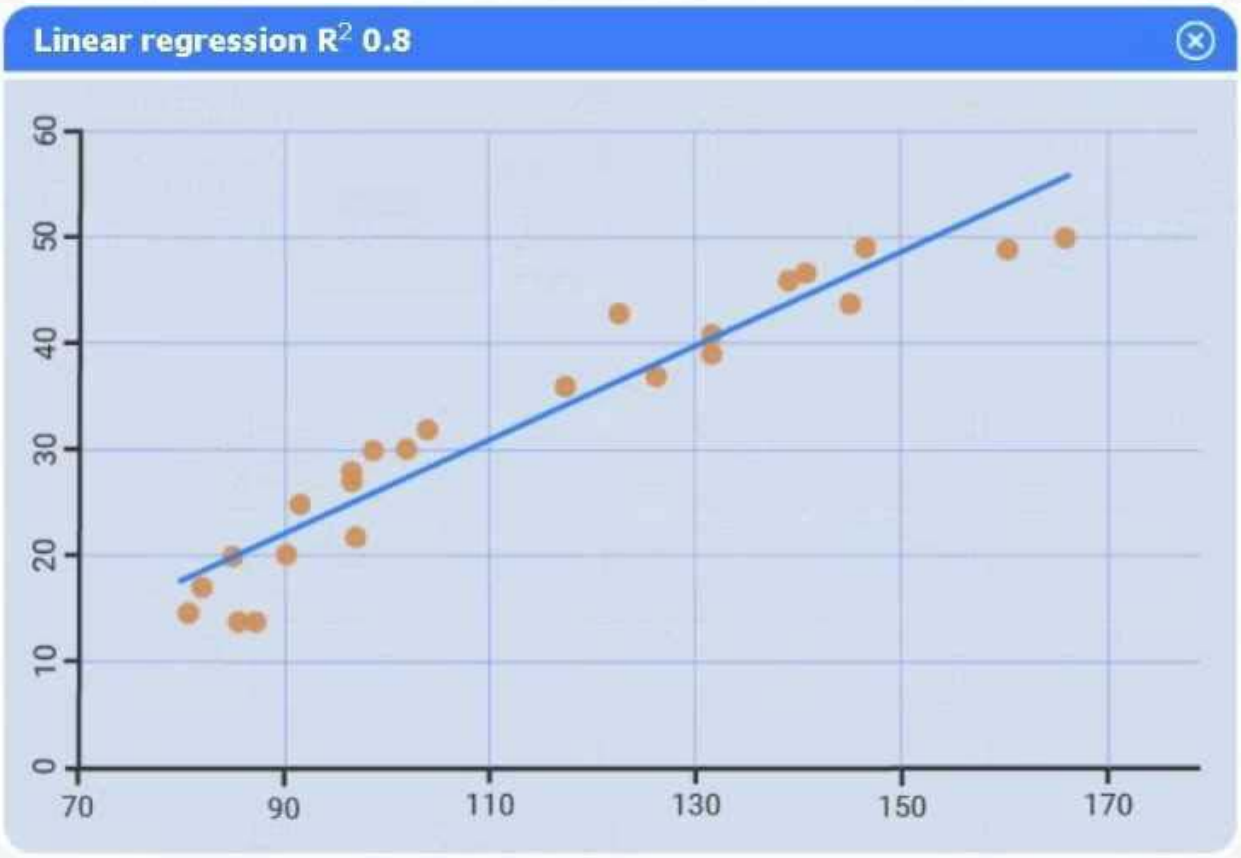
Time	Var 1 Sales (in millions)	Var 2 ROI (% of overall)	R^2 Value
1	3.118026935	6%	
2	4.823728572	11%	
3	7.149131157	18%	
4	2.173859679	5%	
5	3.519662597	9%	
6	5.98246748	12%	
7	8.495414141	14%	
8	3.678906129	7%	
9	3.539605808	6%	

Ridge regression R^2 0.5

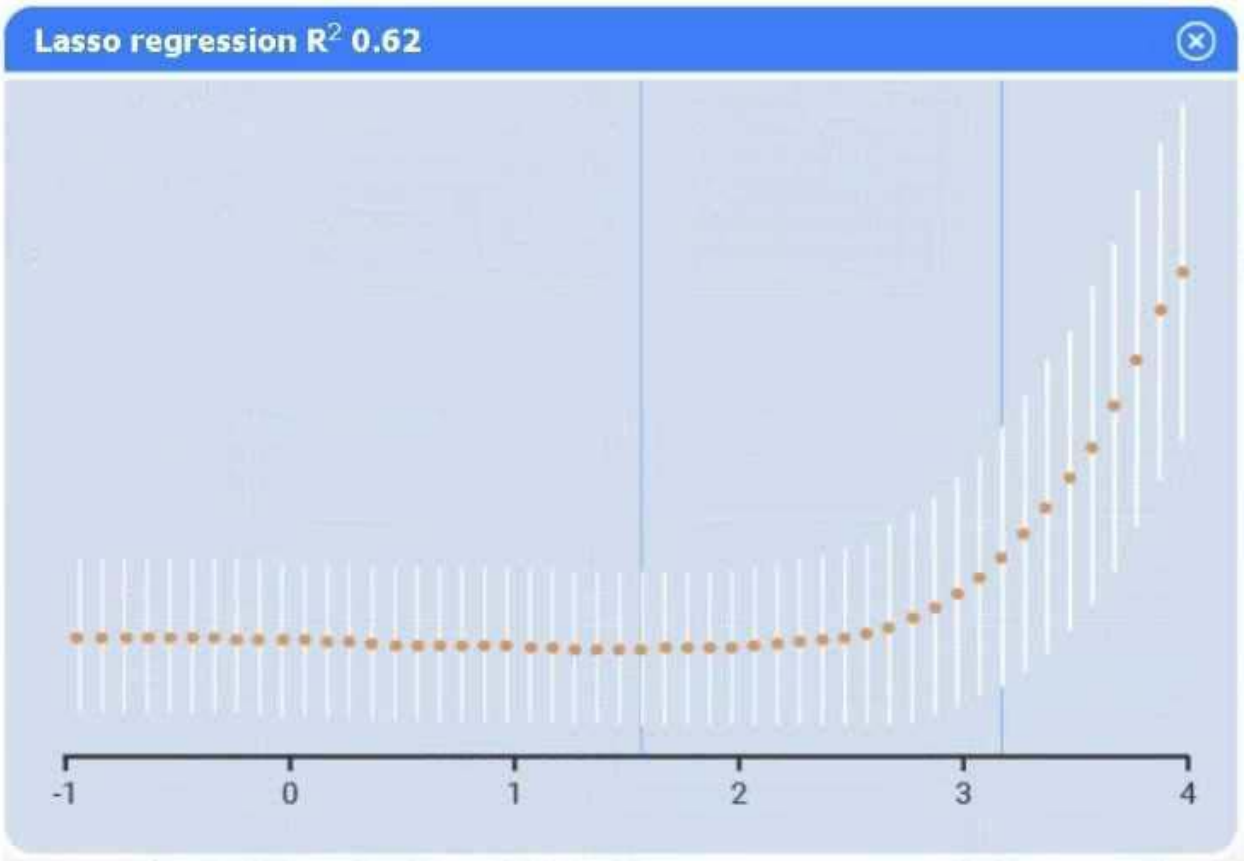


Visit us at: https://www.examsboost.com/test/dy0-001





Visit us at: https://www.examsboost.com/test/dy0-001



Visit us at: https://www.examsboost.com/test/dy0-001

Part 1 | Part 2

Time	Var 1 Sales (in millions)	Var 2 ROI (% of overall)	Var 3 Inventory cost	Var 4 Net operations cost	Var 5 Initial investment
1	326.311584	16%	58	32	24
2	507.9584031	8%	57	50	39
3	232.5685962	5%	53	23	30
4	117.3342091	7%	50	11	35
5	242.866515	7%	60	24	23
6	359.6300247	14%	50	35	38
7	119.384542	19%	56	11	21
8	372.064584	5%	56	37	29
9	320.0212452	18%	51	31	34

View summary output

Which of the following additional variables should the data scientist include in the new model?

- ○ Var 5 Initial investment
- ○ Var 4 Net operations cost
- ○ Var 3 Inventory cost
- ○ None of the variables should be included

Summary output

Regression statistics	
Multiple R	0.999978259
R square	0.999956518
Adjusted R square	0.999913036
Standard error	1.100286825
Observations	9

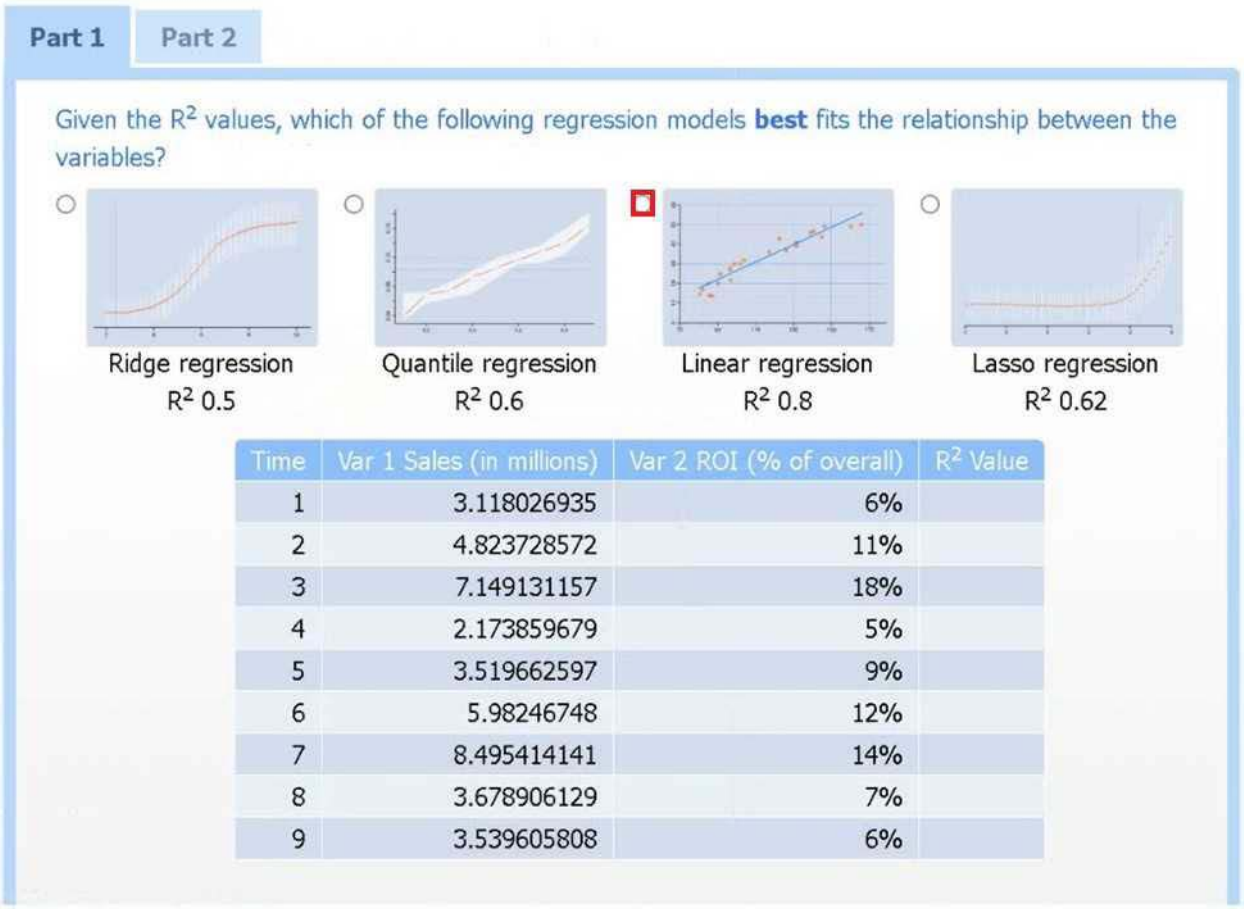
	Coefficients	Standard error	t-stat	p-value
Intercept	30.24229003	9.306229821	3.249682267	0.031385159
Var 2 ROI (% of overall)	50.72139711	13.14967361	3.857236202	0.018190028
Var 3 Inventory cost	-0.315571292	2.013342425	-0.15674	0.89873
Var 4 Net operations cost	9.854244454	0.049842563	197.7074192	0
Var 5 Initial investment	-0.268287655	0.103591751	-1.7654	0.234464

	df	SS	MS	F	Significance F
Regression	4	111363.9712	27840.9928	22997.0904	5.67185E-09
Residual	4	4.842524393	1.210631098		
Total	8	111368.8137			

Answer: See explanation below.

Visit us at: https://www.examsboost.com/test/dy0-001

Explanation:
Part 1
Linear regression.
Of the four models, linear regression has the highest R^2 (0.8), indicating it explains the greatest proportion of variance in sales.



Time	Var 1 Sales (in millions)	Var 2 ROI (% of overall)	R^2 Value
1	3.118026935	6%	
2	4.823728572	11%	
3	7.149131157	18%	
4	2.173859679	5%	
5	3.519662597	9%	
6	5.98246748	12%	
7	8.495414141	14%	
8	3.678906129	7%	
9	3.539605808	6%	

Part 2
Var 4 – Net operations cost.
Net operations cost has a p-value of essentially 0 (far below 0.05), indicating it is the only additional predictor statistically significant in explaining sales. Neither inventory cost ($p \approx 0.90$) nor initial investment ($p \approx 0.23$) reach significance.

Visit us at: https://www.examsboost.com/test/dy0-001

Part 1 | Part 2

Time	Var 1 Sales (in millions)	Var 2 ROI (% of overall)	Var 3 Inventory cost	Var 4 Net operations cost	Var 5 Initial investment
1	326.311584	16%	58	32	24
2	507.9584031	8%	57	50	39
3	232.5685962	5%	53	23	30
4	117.3342091	7%	50	11	35
5	242.866515	7%	60	24	23
6	359.6300247	14%	50	35	38
7	119.384542	19%	56	11	21
8	372.064584	5%	56	37	29
9	320.0212452	18%	51	31	34

View summary output

Which of the following additional variables should the data scientist include in the new model?

- ○ Var 5 Initial investment
- ■ Var 4 Net operations cost
- ○ Var 3 Inventory cost
- ○ None of the variables should be included

Question: 3

A data scientist is building an inferential model with a single predictor variable. A scatter plot of the independent variable against the real-number dependent variable shows a strong relationship between them. The predictor variable is normally distributed with very few outliers. Which of the following algorithms is the best fit for this model, given the data scientist wants the model to be easily interpreted?

A. A logistic regression
B. An exponential regression
C. A linear regression
D. A probit regression

Answer: C

Question: 4

A data scientist wants to evaluate the performance of various nonlinear models. Which of the following is best suited for this task?

A. AIC
B. Chi-squared test
C. MCC
D. ANOVA

Answer: A

Question: 5

Which of the following is the layer that is responsible for the depth in deep learning?

A. Convolution
B. Dropout
C. Pooling
D. Hidden

Answer: D

Thank You for Trying Our Product

For More Information – Visit link below:

https://www.examsboost.com/

15 USD Discount Coupon Code:

G74JA8UF

FEATURES

- ✓ 90 Days Free Updates
- ✓ Money Back Pass Guarantee
- ✓ Instant Download or Email Attachment
- ✓ 24/7 Live Chat Support
- ✓ PDF file could be used at any Platform
- ✓ 50,000 Happy Customer