Answer 1

The difference between Rahul's Training and Testing accuracy can be due to **Overfitting of his model on Train Data**, which is due to High Variance of Model.

Rahul's model must have memorized training data points instead of capturing the Trend or pattern in data resulting in bad performance on unseen data.

Rahul may have built a very complex model using all the predictors and No steps were taken to reduce multi-collinearity, Complexity or Variance of model resulting in a very Over fitted model.

Another reason for overfitting can be use of Higher Degree Polynomial for fitting and thus resulting in a very complex Model.

Rahul's problem can be solved using following steps:

a) Creating a **simple and generic model**.
b) Feature selection using **RFE or manually based on P-Values and VIF**.
c) Using a better metric for checking model performance like **adjusted R-Square, AIC, BIC**, which will **add a penalty term for number of features used.**
d) Using a **LASSO, Ridge or ElasticNet Regression** can also help in reducing model Complexity be reducing the magnitude of Coefficients.

Answer 2

	Ridge Regression	**Lasso Regression**
1	In Ridge Regression the regularization term **sum of the squares of the coefficients** is added to cost function of regression.	In Lasso Regression the regularization term **absolute value of the coefficients** is added to cost function of regression.
2	Ridge regression almost always has a matrix representation for the solution.	Lasso requires iterations to get to the final solution.
3	Ridge regression is computationally less intensive	Lasso regression is computationally more intensive
4	Ridge regression enforces the β coefficients to be lower in magnitude, but it does not enforce them to be zero	Lasso regression enforces the β coefficients to be zero thus resulting in Feature Selection

Answer 3

Given the fact that both the models perform equally well on the test dataset, magnitudes of coefficients are also almost comparable and **number of features is same** in both Ridge and Lasso based Model, we will **go with Ridge regression based model as it is less computationally intensive** and it almost always has a matrix representation for the solution and does not require iteration to achieve at solution resulting in a better performance.

Answer 4

We can ensure a model to be robust and generalizable by making it more simple.

A model can be made simple by using following methods:

a) Feature selection
b) Dimensionality reduction
c) Removing multicollinearity
d) Using a LASSO, Ridge or ElasticNet Regression to reduce model Complexity be reducing the magnitude of Coefficients.
e) Finding the optimal point of Bias Variance tradeoff by fitting a lower dimension Curve
f) outperform complex models when it sees new data.

This simple models are bound to **perform better on unseen datasets** as they have less Variance resulting in more robustness. Simple models not change significantly if the training data points undergo small changes because they are not overfitted on training Data.

Thus the training accuracy may decrease but test accuracy will increase and overall accuracy of model increases if it is simple but not way simpler.

Answer 5

For modeling we will choose the **value of lambda** in Ridge or Lasso such that **test negative mean score is minimum** or **test mean r2_score is maximum. Since Bias Vs Variance tradeoff will be balanced** and model will be **neither Overfit nor Underfit**.

Between Ridge and Lasso model, we will pick the **Lasso model** as we have observed that lasso shrinks the coefficients of lot of variables to nearly 0 in magnitude resulting in **Feature selection** and **simplifying the model even further**.

Thus we will get **a more generalized** at the same time **fairly accurate model, which will perform well on Unseen Data as verified by our model**s test score on Hold-out set.