**Domande Progetto**

**1-Analisi Dataset**

**-Ha senso rimuovere gli outliers? Quale benificio?**

**-Matrice di correlazione: le grandezze fisiche sono molto correlate, è un problema? Andrebbero tolte alcune?**

**-Ha senso utilizzare le interaction per le grandezze altamente correlate?**

**-Il range di prezzo è molto ampio (da 500 a 1000 dollari), ha senso restringerlo in un set di prezzi più vicino per avere delle stime e previsioni più sensate?**

**2- Regressione Lineare**

**-Posso calcolare gli intervalli di confidenza utilizzando lo standard error che mi esce dalla regressione giusto? Avendo tante osservazioni dovrebbe essere uno standard error preciso giusto?**

**3- K-fold**

**4- Subset Selection**

**5- Lasso e Ridge Regression**

**-I nostri coefficienti sono grandi quindi Lasso e Ridge non penalizzano ed il risultato è molto simile alla regressione lineare normale. È un problema?**

**-Restringere il dataset a prezzi simili cambierebbe qualcosa?**

**-Standardizzare i dati cambierebbe qualcosa? In teoria non è già fatto in automatico da Ridge e Lasso?**

**-Aggiungere il termine interazione anche per i modelli non lineari?**

**6-GAM**

**-Come scegliere tra smooth spline e natural spline? Come scegliere il numero di gradi di libertà?**

**7-Regression Tree**

**8-Bagging e Boosting**