**Domande Progetto**

**1-Analisi Dataset**

**-Ha senso utilizzare le interaction term 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?**

**-Ha senso proseguire le analisi lineari con l’interaction term?**

**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? Non mi serve Bootstrap**

**3- K-fold + Polynomials**

**-Ha senso usare K-fold visto che abbiamo un datset cosi grande?**

**-Ha senso usare delle funzioni polinomiali sui regressori dal grado 1 al grado 6?**

**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?**

**-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**

**Come si può aumentare il numero di split dell’albero?**

**8-Bagging e Boosting**