Q: Why do some coefficients switch signs in different linear regression models, e.g. when certain features are dropped?

> This is likely because of multi-collinearity in the data set, i.e. features are highly correlated to one another – e.g. if it is on an island, it is not inland. Collinearity can make coefficient estimates highly unstable. Linear regression is especially vulnerable to this problem. Note collinearity affects how you can interpret the coefficients and *p*-values but not the predictions and the fit, so it is relevant only when interpreting the model.
>
> Tree-based models (random forest) handle correlated variables much better than linear regression. For example, if we have two identical columns, decision tree / random forest will automatically "drop" one of them when splitting.
>
> There is a very informative explanation of all the potential problems in a linear regression model in my favourite statistics book (*Introduction to Statistical Learning with Applications in R*), which can be accessed here: https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf. See Section 3.3.3 (page 92) for the six potential problems in linear regression (non-linearity of the response-predictor relationship, correlation of error terms, non-constant variance of error terms, outliers, high-leverage points, and collinearity). See page 99-102 for an illustration of the concept of collinearity.

Q: When you get the feature importances, how do you know how they are ordered?

> The order is the same as the order in the dataset. So the column names are just the X.columns.

Q: Under what circumstances would we use MinMax Scaler over the Standard Scaler?

> MinMax Scaler transforms each value in the column proportionally within the range [0,1] and will preserve the shape of the data set (no distortion). StandardScaler should be used primarily for normally distributed data.

Q: What scaler is best if there are outliers?

> You can use the RobustScaler to reduce the influence of the outliers or alternatively, you can remove / impute the outliers.

Q: Why did we use stratified sampling on income rather than housing value?

> You use stratified sampling if values are not evenly distributed in the dataset to give your algorithm enough evidence for each range (stratum). Since we are trying to predict housing value, we can't stratify by the outcome we are trying to predict since that risks over-fitting to the stratification method. We can only stratify based on input features.

Q: Why would you use randomized search over grid search?

If there is a large feature set, Randomized Search selects “random” combinations and is quicker and more appropriate. Grid Search performs exhaustive search over the hyperparameter combinations and would be slower in large data sets – it doesn't make a huge difference in this practical due to the small number of features.