1. Import numpy, pandas, seaborn, matplotlib
2. Import data as “training” and “X\_test”, print “training” ‘s columns, shape. Print X\_test shape
   * Check training: info
3. Write function to create scrollable table “create\_scrollable\_table”
4. Create numerical dataset as “df\_num”. Use “create\_scrollable\_table” on statistics summary of df\_num (transpose) 🡺 display
   * Check df\_num: info, shape
5. Create categorical dataset as “df\_cat”. Use “create\_scrollable\_table” on statistics summary of df\_cat (transpose) 🡺 display
   * Check df\_cat: info, shape
6. Check null values on each column. Use “create\_scrollable\_table” to display
7. Check percentage of null values on each column. Use “create\_scrollable\_table” to display
   * Check those with null percentage > 5%
8. Show rows with missing values. Use “create\_scrollable\_table” to display
9. Create a “SalePrice Distribution” graph, consisting of:
   * A histogram of “SalePrice”
   * A normal distribution line
10. Create a “Q-Q plot”, consisting of:
    * A Q-Q plot
    * Best line fit
11. Write function to show value on bar graphs.
12. Make bar graphs showing:
    * Distribution of building types
    * Distribution of dwelling types and their relation to average sale prices
13. Make bar graphs showing the impact of zoning on sale prices.
14. Make bar graphs showing the impact of street and alley on sale prices.
15. Make bar graphs showing the
    * Average sale price by shape.
    * Average sale price by contour.
16. Make a scatter graph showing the relationship between SalePrice and PropertyAge (needs calculation)
17. Make a scatter graph showing the relationship between SalePrice and Living Area
    * Also, calculate the correlation between SalePrice and Living Area
18. Make a box plot showing SalePrice trends over the years.
19. Import: ColumnTransformer, Pipeline, SimpleImputer, StandardScaler, OneHotEncoder
20. Define transformers for numerical and categorical columns.
21. Use ColumnTransformer to apply numerical\_transformer for numerical columns and categorical\_transformer for categorical columns.
22. Create preprocessor Pipeline. Create X\_train, y\_train. Apply pipeline on those training sets.
23. Import: LinearRegression, RandomForestRegressor, XGBRegressor, GridSearchCV, KFold
24. Make a “models” dictionary containing LinearRegression, RandomForestRegressor, XGBRegressor
25. Define param\_grids. Define cv=KFold
26. Run GridSearchCV/ Fit training datas on each mode. Print best\_params and best\_score of each model.
27. Import PCA