Project Statement: House Pricing Regression for Cuckoo Cribs Corporation

Note: Cuckoo Cribs Corporation is such a dysfunctional corporation they can't even write their own name property on the company's logo.

Welcome team! Our newest venture at Cuckoo Cribs Corporation involves predicting house prices using regression techniques and presenting a model to the board of directors with the "pricing machine". Through analysis and modelling, we aim to convince the board that decision-making processes related to property pricing should be done using machine learning and not via the colour of the shirt that the CEO is using on a particular day.

To achieve that, we need to convince them that **data-driven decision making is the way to go. We will build a model using KNIME Analytics and show them that we can achieve good performance in predicting housing prices accurately.**

With this machine learning model, the team at Cuckoo Cribs will want to find the "intrinsic value" of a property, finding good deals on the market, fast. To show them how they can use data to make decisions, you should provide a report (with no more than 5 or 6 pages, excluding annex) where you need to address the following points:

1. Executive Summary:

Our objective is to provide insights into the existing situation, model performance, business outcomes, deployment strategies, and suggestions for improvement, all tailored for our board, who might not be the savviest when it comes to data science.

2. Available Data:

Type of data used, feature engineering or other topics related to the available dataset.

3. Model Details and Performance:

Describe the type of model you've used and the technical results. Don't forget that Cuckoo Cribs' board is very old fashioned and you need to do a compelling argument about the type of model you've used.

4. Business Outcome:

What's the impact of deploying that model? Can you creatively estimate some scenarios of the real impact in Cuckoo Cribs business?

5. Deployment:

How would you deploy this model in real life? Describe some ideas of how real estate agents would use the model you've produced in real life.

6. Monitoring and Retraining:

Regular monitoring of the model's performance is essential. What's your recommendation for retraining the model and keep track the relationship of the features and target is still relevant?

7. Improvements and Next Steps:

Describe some recommendations of next steps for future model developments.

9. Annexes:

Supplementary materials such as technical documentation, code snippets, and additional analyses will be included for reference.

Deliverables of the project:

- KNIME Workflow with trained regression model. Make sure that every KNIME workflow's nodes run with the dataset realestatedata.csv
- Report with the aforementioned structure (max 6 pages without annexes).
- Predictions for the test set (unseen data) given by the teachers.
- When detailing the error metric of your project, please describe RMSE and MAPE.

As mentioned, the test set is some data set on the side, where teachers can access additional ground truth examples. Your submission will be compared against those examples and a leaderboard will be created based on the **RMSE** error.

Column Metadata:

SalePrice - the property's sale price in dollars. This is the target variable that you're trying to predict.
MSSubClass: The building class
MSZoning: The general zoning classification
LotFrontage: Linear feet of street connected to property
LotArea: Lot size in square feet
Street: Type of road access
Alley: Type of alley access
LotShape: General shape of property
LandContour: Flatness of the property
Utilities: Type of utilities available
LotConfig: Lot configuration
LandSlope: Slope of property
Neighborhood: Physical locations within city limits
Condition1: Proximity to main road or railroad
Condition2: Proximity to main road or railroad (if a second is present)
BldgType: Type of dwelling
HouseStyle: Style of dwelling
OverallQual: Overall material and finish quality
OverallCond: Overall condition rating
YearBuilt: Original construction date
YearRemodAdd: Remodel date
RoofStyle: Type of roof
RoofMatl: Roof material
Exterior1st: Exterior covering on house
Exterior2nd: Exterior covering on house (if more than one material)
MasVnrType: Masonry veneer type
MasVnrArea: Masonry veneer area in square feet
ExterQual: Exterior material quality
ExterCond: Present condition of the material on the exterior
Foundation: Type of foundation
BsmtQual: Height of the basement
BsmtCond: General condition of the basement
BsmtExposure: Walkout or garden level basement walls
BsmtFinType1: Quality of basement finished area
BsmtFinSF1: Type 1 finished square feet
BsmtFinType2: Quality of second finished area (if present)
BsmtFinSF2: Type 2 finished square feet
BsmtUnfSF: Unfinished square feet of basement area
TotalBsmtSF: Total square feet of basement area
Heating: Type of heating
HeatingQC: Heating quality and condition
CentralAir: Central air conditioning
Electrical: Electrical system
1stFlrSF: First Floor square feet

2ndFlrSF: Second floor square feet
LowQualFinSF: Low quality finished square feet (all floors)
GrLivArea: Above grade (ground) living area square feet
BsmtFullBath: Basement full bathrooms
BsmtHalfBath: Basement half bathrooms
FullBath: Full bathrooms above grade
HalfBath: Half baths above grade
Bedroom: Number of bedrooms above basement level
Kitchen: Number of kitchens
KitchenQual: Kitchen quality
TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)
Functional: Home functionality rating
Fireplaces: Number of fireplaces
FireplaceQu: Fireplace quality
GarageType: Garage location
GarageYrBlt: Year garage was built
GarageFinish: Interior finish of the garage
GarageCars: Size of garage in car capacity
GarageArea: Size of garage in square feet
GarageQual: Garage quality
GarageCond: Garage condition
PavedDrive: Paved driveway
WoodDeckSF: Wood deck area in square feet
OpenPorchSF: Open porch area in square feet
EnclosedPorch: Enclosed porch area in square feet
3SsnPorch: Three season porch area in square feet
ScreenPorch: Screen porch area in square feet
PoolArea: Pool area in square feet
PoolQC: Pool quality
Fence: Fence quality
MiscFeature: Miscellaneous feature not covered in other categories
MiscVal: $Value of miscellaneous feature
MoSold: Month Sold
YrSold: Year Sold
SaleType: Type of sale
SaleCondition: Condition of sale