



## limpieza de datos

- Quitar columna ID
- Corte, color y claridad: pasar a categóricas ordinales.
  - Diccionario vs. get dummies
- Drop: x, y, z.
- Sin normalización/estandarización



## aplicación de modelos

- **Support Vector Regression** (*RMSE: 4233.30*)
- **Random Forest Regressor** (*RMSE: 543.89*)
  - `rf_reg = RandomForestRegressor(n_estimators=200, max_depth=20, min_samples_leaf=3, random_state=111)`
  - `selector = RFECV(rf_reg, step=1, cv=5)`
- **Gradient Boosting Regressor** (*RMSE: 536.57*)
  - `{'n_estimators': 800,`
  - `'max_depth': 8,`
  - `'min_samples_split': 25,`
  - `'min_samples_leaf': 10`
  - `'learning_rate': 0.01,` #Para evitar el over-fitting. No en random forests.
  - `'loss': 'ls'}`

### Justificación

- Tree-based models:
  - Variables categóricas
  - Missing values
  - Relaciones no-lineales