# Tracking ML Experiments



Paweł Kordek Software Engineer – Data and Machine Learning

@pawel\_kordek



#### Overview



**Dataset** 

Model training pipeline

Vocabulary

**MLflow Tracking: Demo** 



## Dataset



Task

Predict market price of a property given some of its characteristics



### Ames Housing Dataset



De Cock D. 2011. Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project. Journal of Statistics Education; 19(3).



Complete dataset documentation: http://jse.amstat.org/v19n3/decock/DataDocumentation.txt

## Ames Housing Dataset

#### 2930 rows and 83 columns

#### A lot of features!

#### Selected features:

- 'Lot Area' size of the lot in feet
- 'Gr Liv Area' living area above the ground level
- 'Garage Area' garage size in square feet
- 'Bldg Type' building type, e.g. '1Fam'

Target: 'SalePrice' - in USD



Lot Area	Gr Living Area	Garage Area	Bldg Type	SalePrice
31770	1656	528.0	1Fam	215000
11622	896	730.0	1Fam	105000
14267	1329	312.0	1Fam	172000

## One-hot Encoding

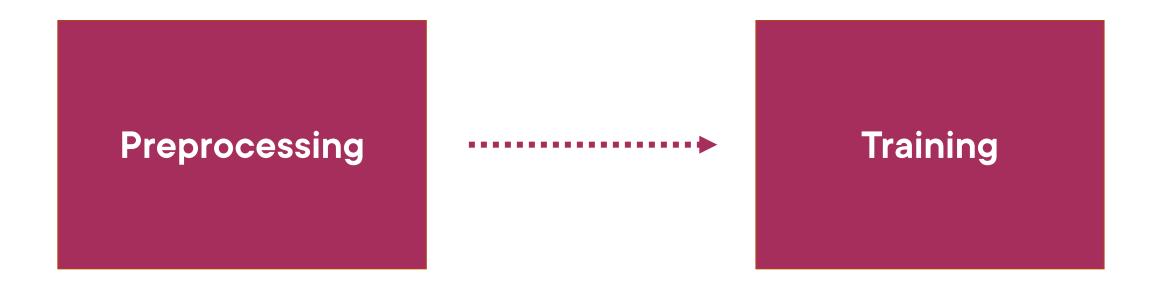
	BldgType
0	1Fam
1	Duplx
2	Twnhsl
3	Duplx



	1Fam	Duplx	Twnhsl
0	1	0	0
1	0	1	0
2	0	0	1
3	0	1	0



# Experimenting with ML Models





Preprocessing (One-hot encoding)

Training (Linear regression)

Evaluation (Mean squared error)

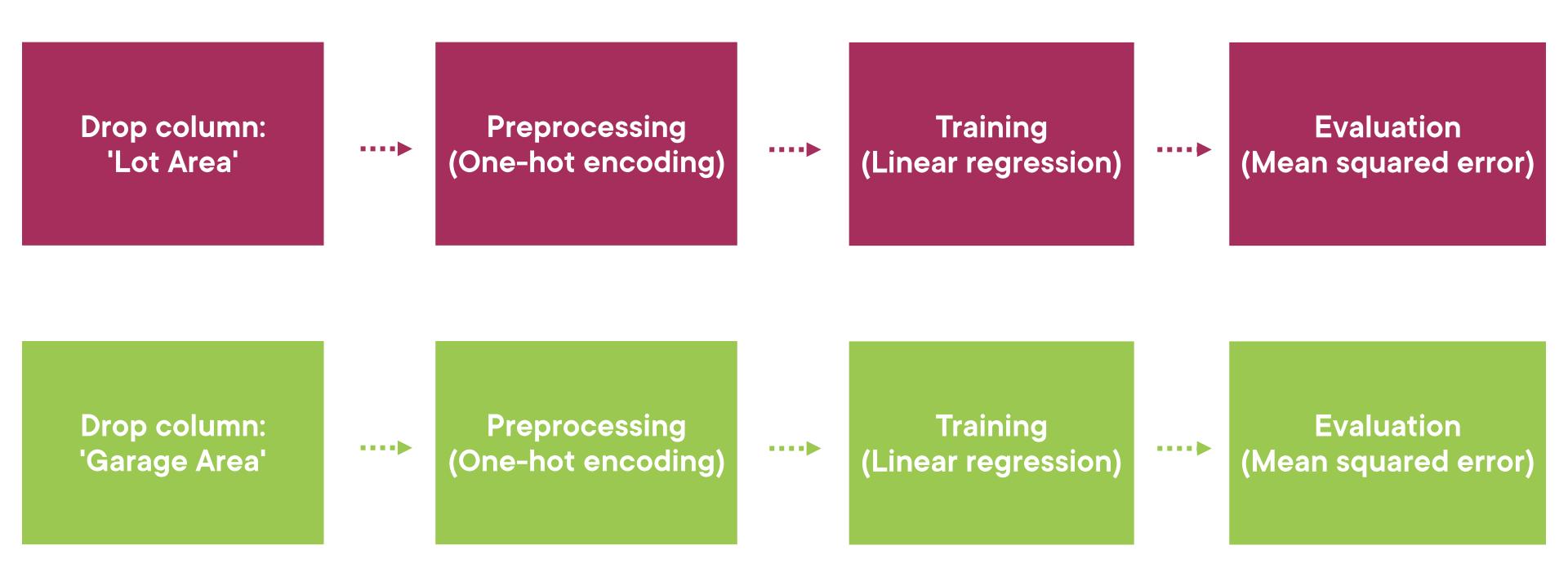
Drop column

Preprocessing (One-hot encoding)

Clinear regression)

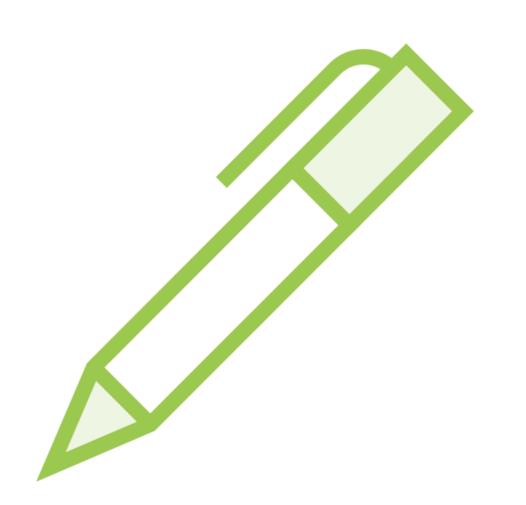
Evaluation (Mean squared error)





### Repeat for all feature columns





Initial training – error: 123.65

Dropped column: Lot Area – error: 154.34

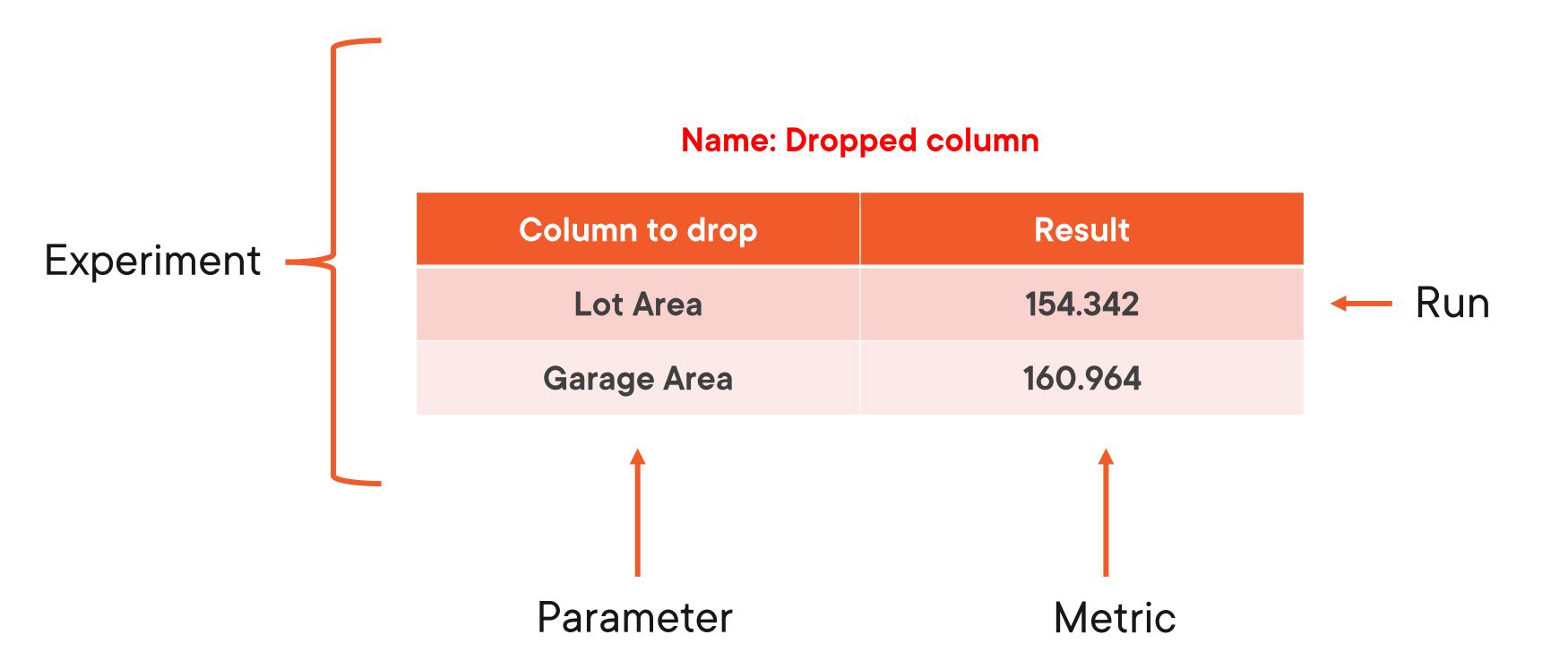
Dropped column: Garage Area – error: 160.96



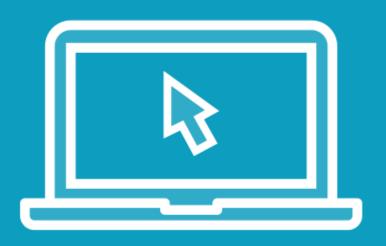
Name	Column to drop	Result
Initial training	_	123.654
Dropped column	Lot area	154.342
Dropped column	Garage area	160.964

Name	Result
Initial training	123.654

Name	Column to drop	Result
Dropped column	Lot Area	154.342
Dropped column	Garage Area	160.964



### Demo



Simple ML model for house price predictions

Creating MLflow experiments and runs

Logging parameters and metrics

**Using the MLflow UI** 



### Summary



#### **Dataset**

#### **Training process**

#### Concepts

- Experiment
- Run
- Parameter
- Metric

**MLflow Python library** 

**MLflow UI** 

