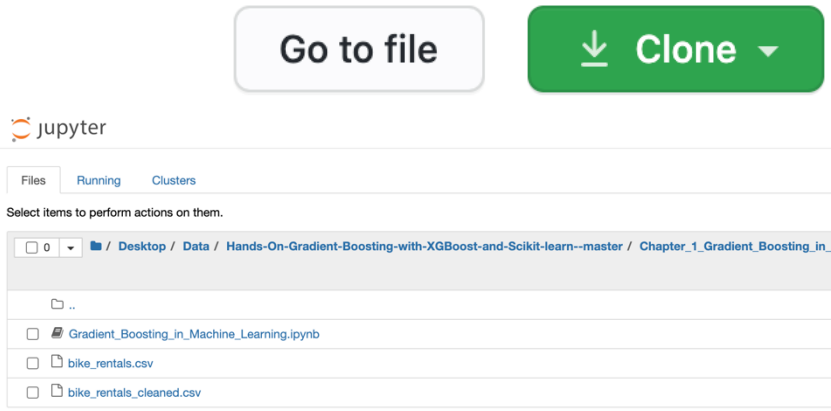
Chapter 1: Machine Learning Landscape




```
# Import pandas
import pandas as pd
```

```
# Upload 'bike_rentals.csv' to dataFrame
df_bikes = pd.read_csv('bike_rentals.csv')
```

```
# Display first 5 rows
df_bikes.head()
```

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
0	1	2011-01-01	1.0	0.0	1.0	0.0	6.0	0.0	2	0.344167	0.363625	0.805833	0.160446	331	654	985
1	2	2011-01-02	1.0	0.0	1.0	0.0	0.0	0.0	2	0.363478	0.353739	0.696087	0.248539	131	670	801
2	3	2011-01-03	1.0	0.0	1.0	0.0	1.0	1.0	1	0.196364	0.189405	0.437273	0.248309	120	1229	1349
3	4	2011-01-04	1.0	0.0	1.0	0.0	2.0	1.0	1	0.200000	0.212122	0.590435	0.160296	108	1454	1562
4	5	2011-01-05	1.0	0.0	1.0	0.0	3.0	1.0	1	0.226957	0.229270	0.436957	0.186900	82	1518	1600

	instant	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed
count	731.000000	731.000000	730.000000	730.000000	731.000000	731.000000	731.000000	731.000000	730.000000	730.000000	728.000000	726.000000
mean	366.000000	2.496580	0.500000	6.512329	0.028728	2.997264	0.682627	1.395349	0.495587	0.474512	0.627987	0.190476
std	211.165812	1.110807	0.500343	3.448303	0.167155	2.004787	0.465773	0.544894	0.183094	0.163017	0.142331	0.077725
min	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000	0.059130	0.079070	0.000000	0.022392
25%	183.500000	2.000000	0.000000	4.000000	0.000000	1.000000	0.000000	1.000000	0.336875	0.337794	0.521562	0.134494
50%	366.000000	3.000000	0.500000	7.000000	0.000000	3.000000	1.000000	1.000000	0.499167	0.487364	0.627083	0.180971
75%	548.500000	3.000000	1.000000	9.750000	0.000000	5.000000	1.000000	2.000000	0.655625	0.608916	0.730104	0.233218
max	731.000000	4.000000	1.000000	12.000000	1.000000	6.000000	1.000000	3.000000	0.861667	0.840896	0.972500	0.507463

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
56	57	2011-02-26	1.0	0.0	2.0	0.0	6.0	0.0	1	0.282500	0.282192	0.537917	NaN	424	1545	1969
81	82	2011-03-23	2.0	0.0	3.0	0.0	3.0	1.0	2	0.346957	0.337939	0.839565	NaN	203	1918	2121
128	129	2011-05-09	2.0	0.0	5.0	0.0	1.0	1.0	1	0.532500	0.525246	0.588750	NaN	664	3698	4362
129	130	2011-05-10	2.0	0.0	5.0	0.0	2.0	1.0	1	0.532500	0.522721	NaN	0.115671	694	4109	4803
213	214	2011-08-02	3.0	0.0	8.0	0.0	2.0	1.0	1	0.783333	0.707071	NaN	0.205850	801	4044	4845
298	299	2011-10-26	4.0	0.0	10.0	0.0	3.0	1.0	2	0.484167	0.472846	0.720417	NaN	404	3490	3894
388	389	2012-01-24	1.0	1.0	1.0	0.0	2.0	1.0	1	0.342500	0.349108	NaN	0.123767	439	3900	4339
528	529	2012-06-12	2.0	1.0	6.0	0.0	2.0	1.0	2	0.653333	0.597875	0.833333	NaN	477	4495	4972
701	702	2012-12-02	4.0	1.0	12.0	0.0	0.0	0.0	2	NaN	NaN	0.823333	0.124379	892	3757	4649
730	731	2012-12-31	1.0	NaN	NaN	0.0	1.0	0.0	2	0.215833	0.223487	0.577500	0.154846	439	2290	2729

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
56	57	2011-02-26	1.0	0.0	2.0	0.0	6.0	0.0	1	0.282500	0.282192	0.537917	0.180971	424	1545	1969
81	82	2011-03-23	2.0	0.0	3.0	0.0	3.0	1.0	2	0.346957	0.337939	0.839565	0.180971	203	1918	2121

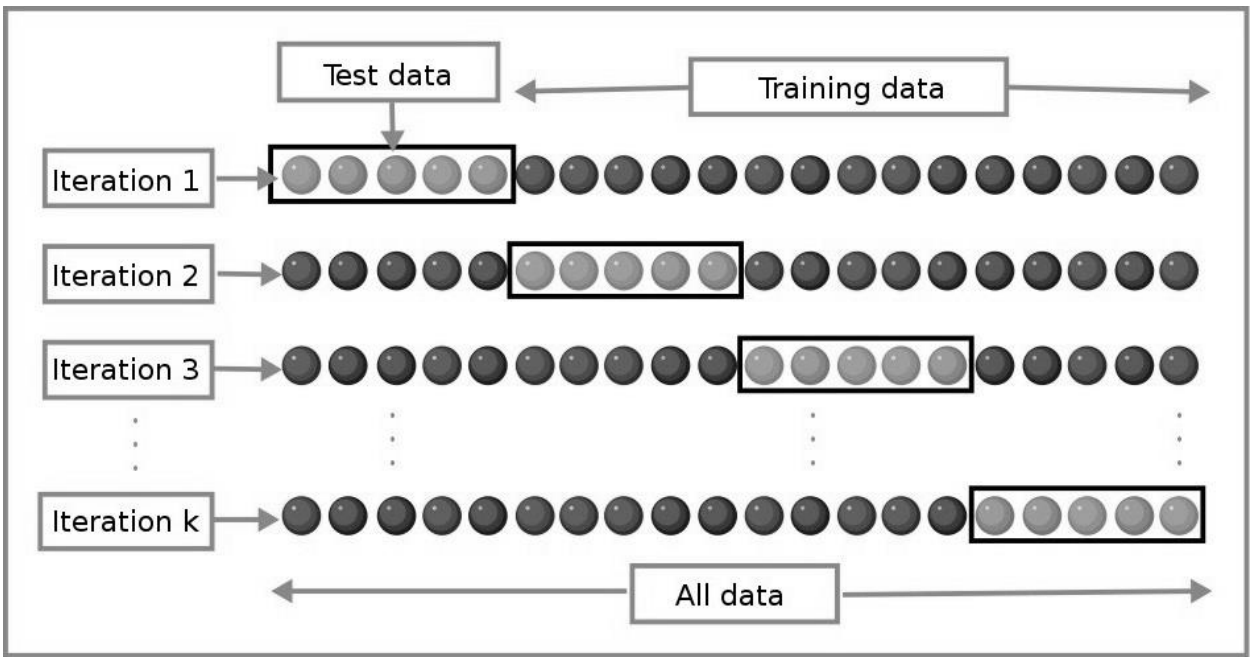
season	instant	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
1.0	366.0	0.5	2.0	0.0	3.0	1.0	1.0	0.285833	0.282821	0.543750	0.202750	218.0	1867.0	2209.0
2.0	308.5	0.5	5.0	0.0	3.0	1.0	1.0	0.562083	0.538212	0.646667	0.191546	867.0	3844.0	4941.5
3.0	401.5	0.5	8.0	0.0	3.0	1.0	1.0	0.714583	0.656575	0.635833	0.165115	1050.5	4110.5	5353.5
4.0	493.0	0.5	11.0	0.0	3.0	1.0	1.0	0.410000	0.409708	0.661042	0.167918	544.5	3815.0	4634.5

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
701	702	2012-12-02	4.0	1.0	12.0	0.0	0.0	0.0	2	NaN	NaN	0.823333	0.124379	892	3757	4649

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
726	727	2012-12-27	1.0	1.0	12	0.0	4.0	1.0	2	0.254167	0.226642	0.652917	0.350133	247	1867	2114
727	728	2012-12-28	1.0	1.0	12	0.0	5.0	1.0	2	0.253333	0.255046	0.590000	0.155471	644	2451	3095
728	729	2012-12-29	1.0	1.0	12	0.0	6.0	0.0	2	0.253333	0.242400	0.752917	0.124383	159	1182	1341
729	730	2012-12-30	1.0	1.0	12	0.0	0.0	0.0	1	0.255833	0.231700	0.483333	0.350754	364	1432	1796
730	731	2012-12-31	1.0	NaN	12	0.0	1.0	0.0	2	0.215833	0.223487	0.577500	0.154846	439	2290	2729

```
# Initialize LinearRegression model
lin_reg = LinearRegression()

# Fit lin_reg on training data
lin_reg.fit(X_train, y_train)

# Predict X_test using lin_reg
y_pred = lin_reg.predict(X_test)

# Import mean_squared_error
from sklearn.metrics import mean_squared_error

# Import numpy
import numpy as np

# Compute mean_squared_error as mse
mse = mean_squared_error(y_test, y_pred)

# Compute root mean squared error as rmse
rmse = np.sqrt(mse)

# Display root mean squared error
print("RMSE: %0.2f" % (rmse))
```

```
RMSE: 898.21
```



	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
0	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
1	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
2	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
3	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
4	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K

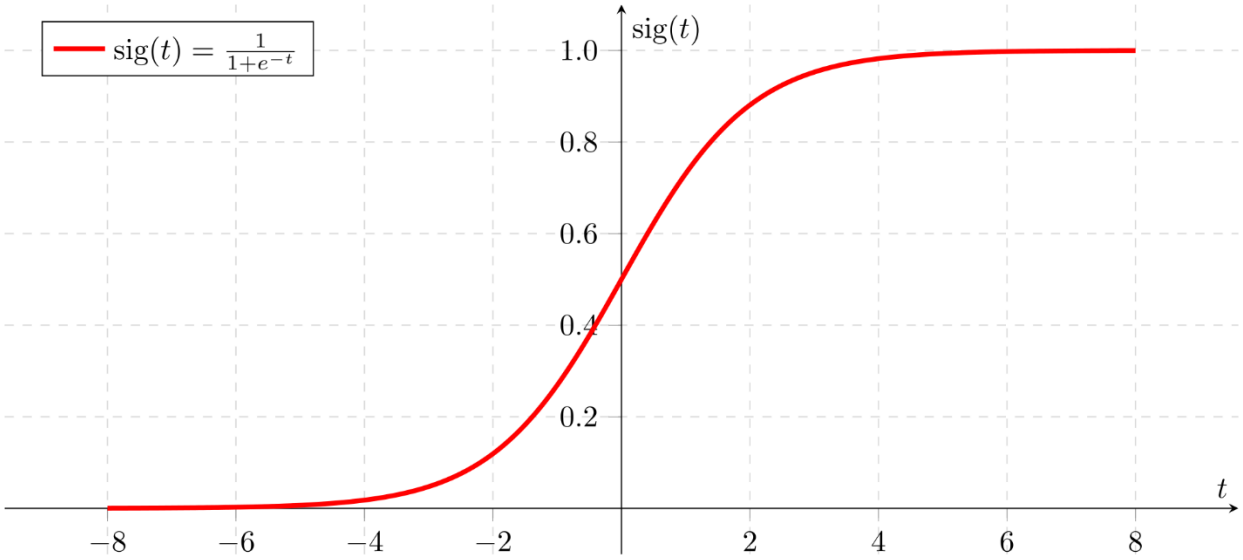
	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	income
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K

	Book Types
0	hardback
1	paperback
2	ebook

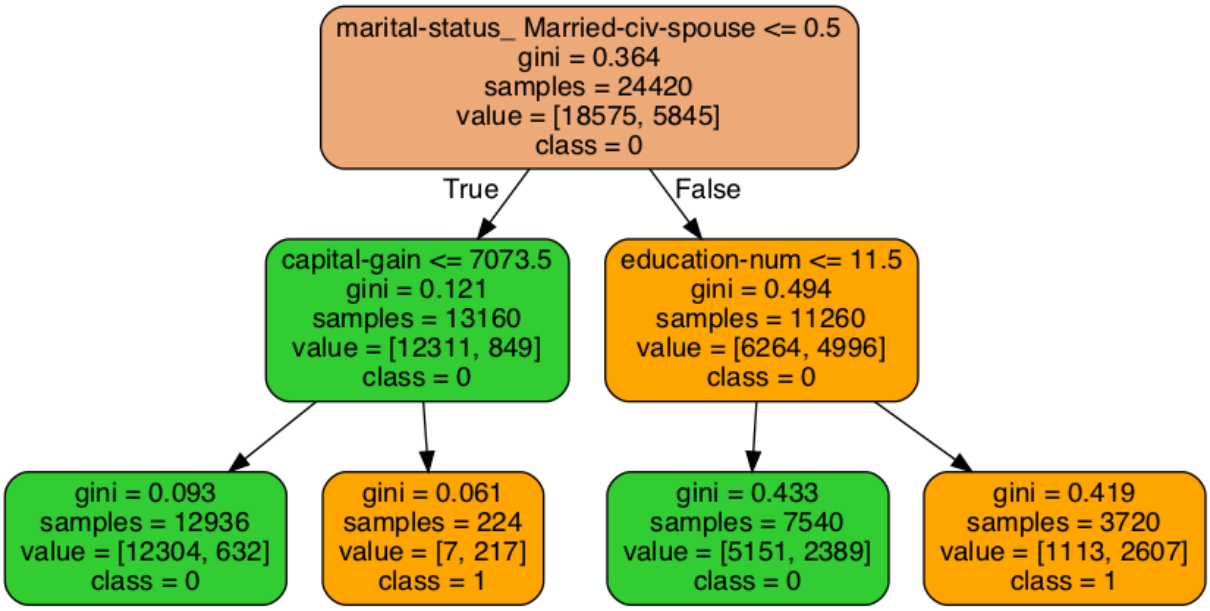
	hardback	paperback	ebook
0	1	0	0
1	0	1	0
2	0	0	1

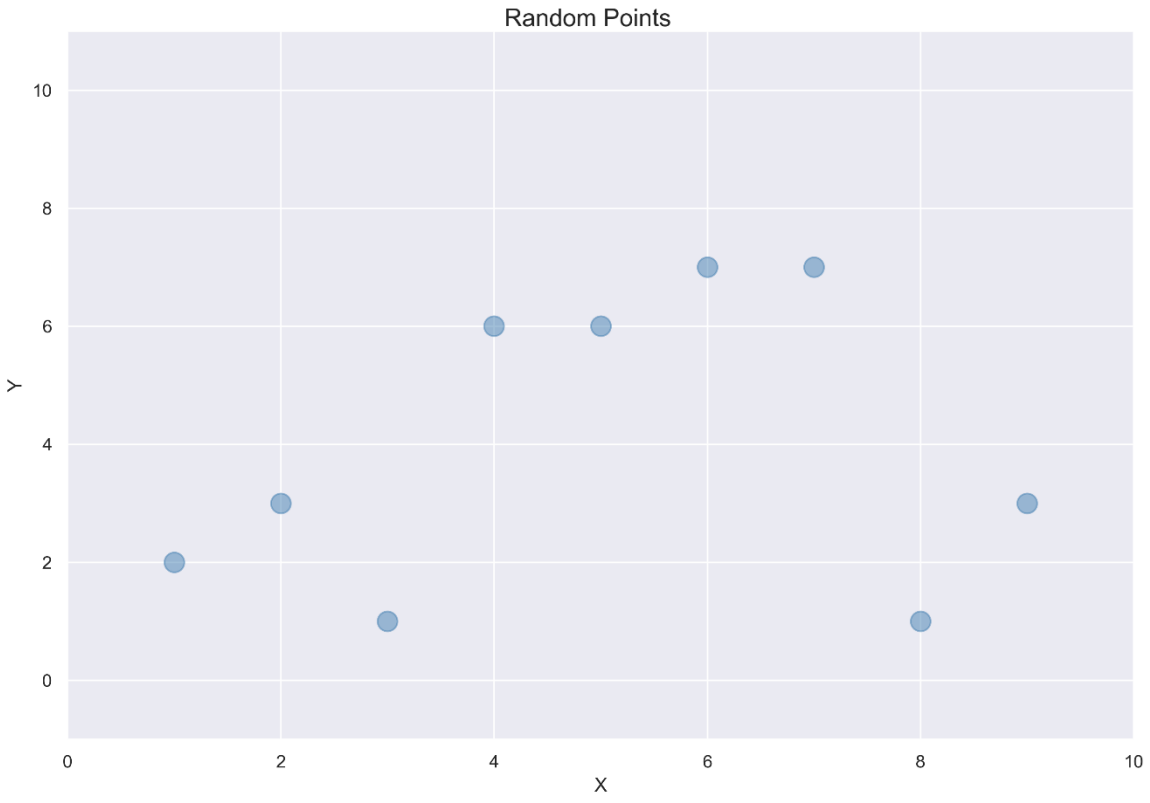
	age	fnlwgt	education-num	capital-gain	capital-loss	hours-per-week	workclass_?	workclass_Federal-gov	workclass_Local-gov	workclass_Never-worked	...	native-country_Scotland	native-country_South	native-country_Taiwan	native-country_Thailand	native-cc Trinadad&T
0	39	77516	13	2174	0	40	0	0	0	0	...	0	0	0	0	
1	50	83311	13	0	0	13	0	0	0	0	...	0	0	0	0	
2	38	215646	9	0	0	40	0	0	0	0	...	0	0	0	0	
3	53	234721	7	0	0	40	0	0	0	0	...	0	0	0	0	
4	28	338409	13	0	0	40	0	0	0	0	...	0	0	0	0	

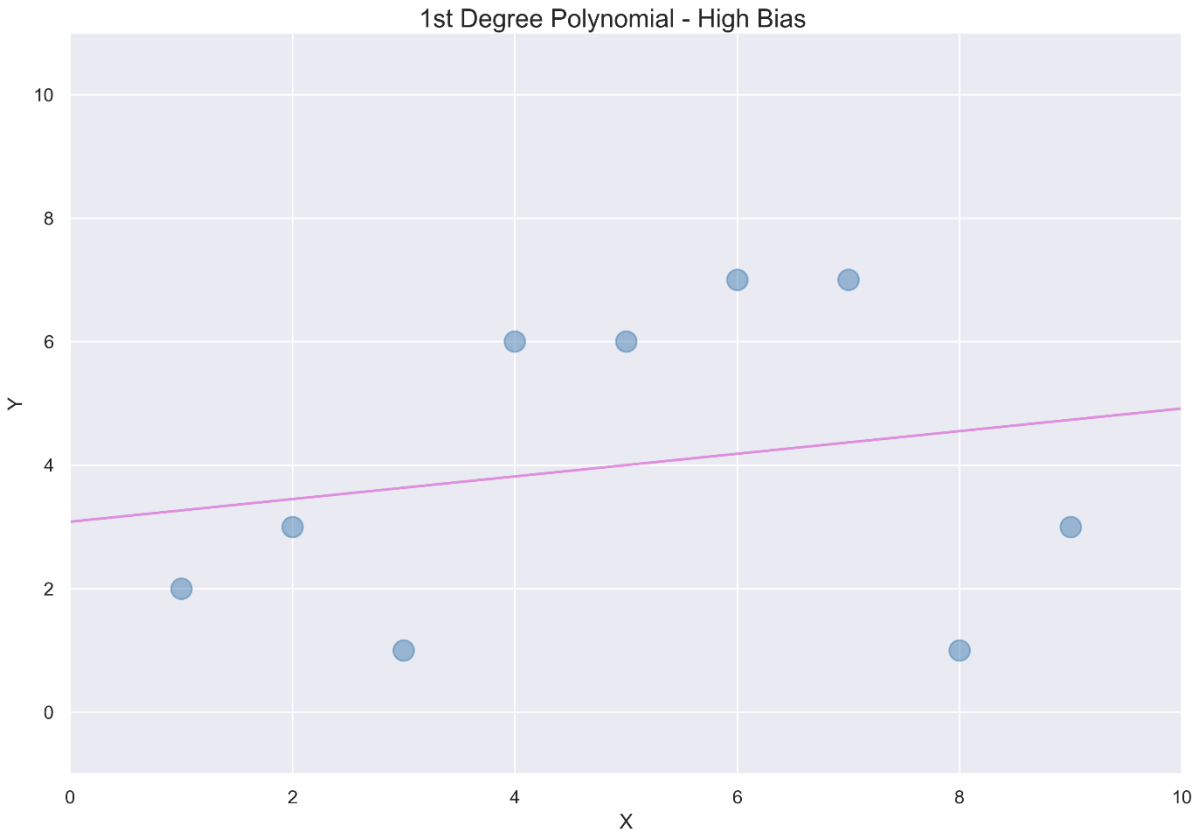
5 rows × 94 columns

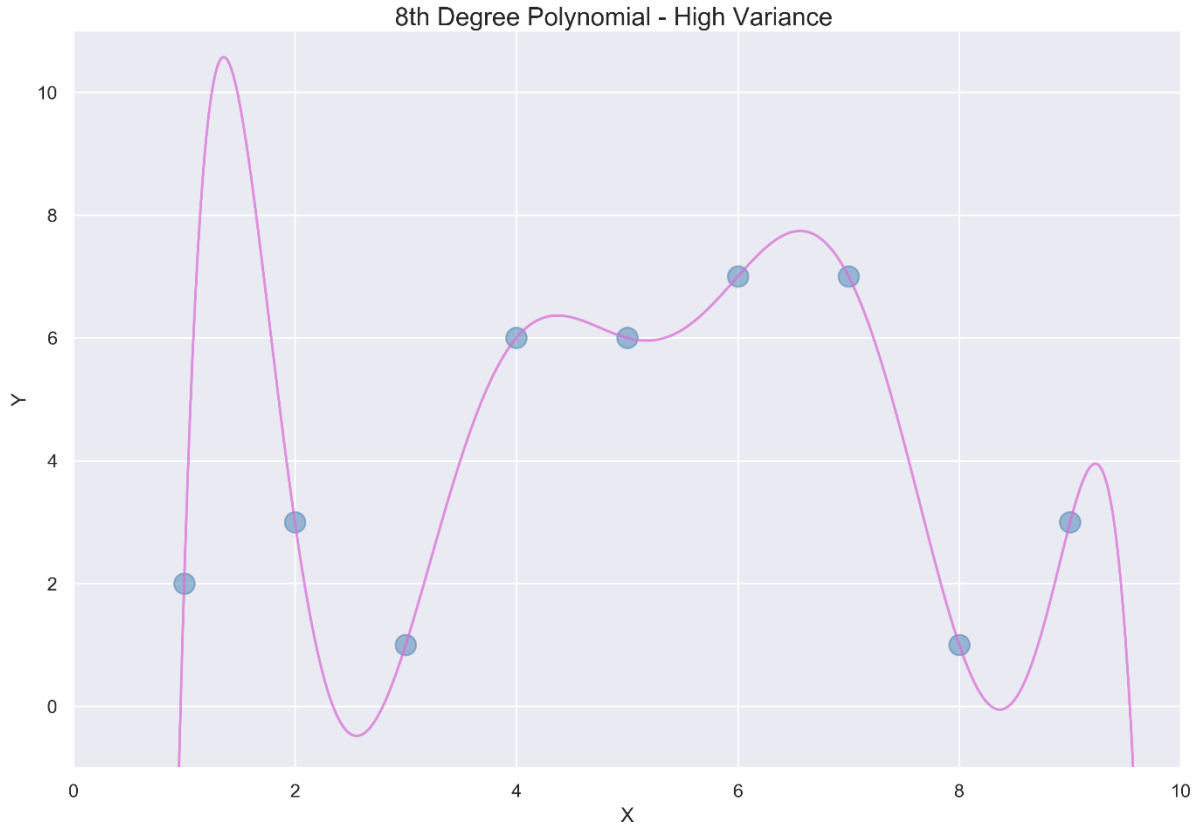


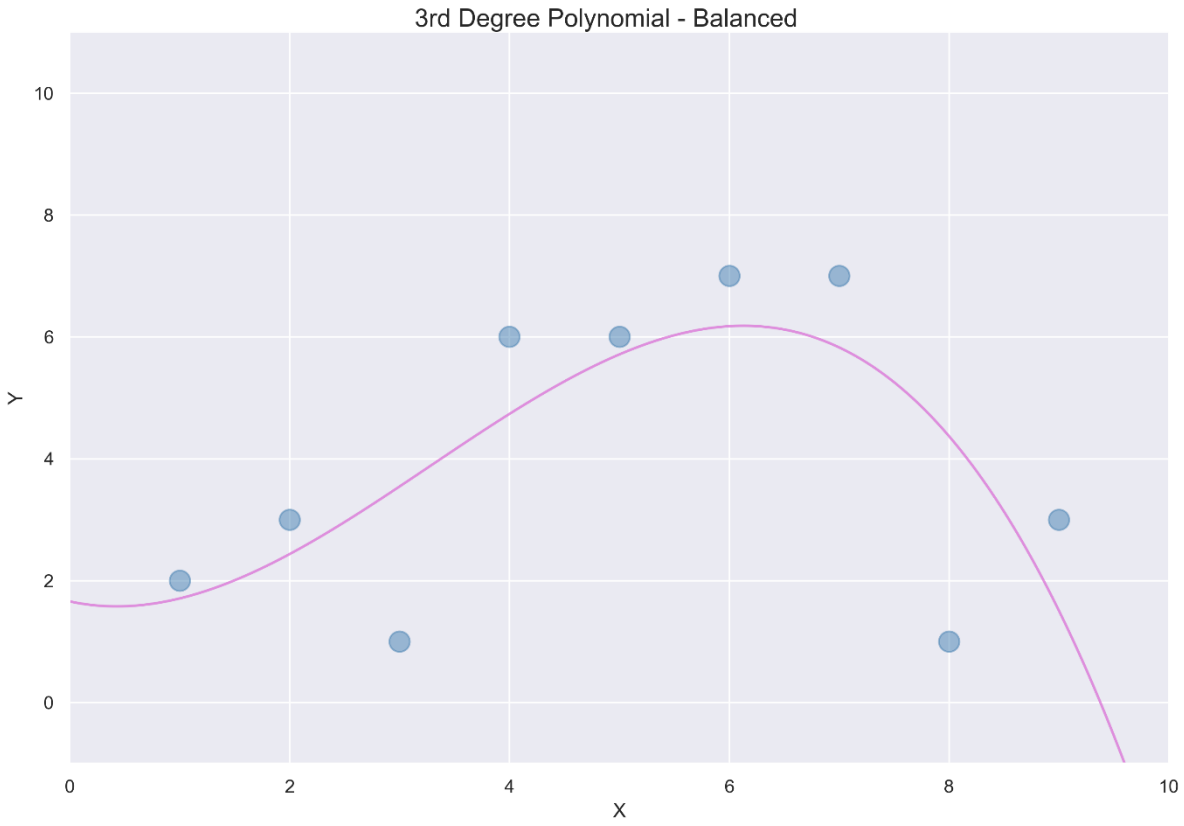
Chapter 2: Decision Trees in Depth






1st Degree Polynomial - High Bias
Y
X
0
2
4
6
8
10


8th Degree Polynomial - High Variance
Y
X
0
2
4
6
8
10


3rd Degree Polynomial - Balanced
10
8
6
4
2
0
Y
0
2
4
6
8
10
X

sklearn.tree.DecisionTreeRegressor

class `sklearn.tree.` **`DecisionTreeRegressor`**(*criterion='mse', splitter='best', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, presort='deprecated', ccp_alpha=0.0*) [source]

A decision tree regressor.

Read more in the User Guide.

Parameters:

criterion : *{"mse", "friedman_mse", "mae"}, default="mse"*
The function to measure the quality of a split. Supported criteria are "mse" for the mean squared error, which is equal to variance reduction as feature selection criterion and minimizes the L2 loss using the mean of each terminal node, "friedman_mse", which uses mean squared error with Friedman's improvement score for potential splits, and "mae" for the mean absolute error, which minimizes the L1 loss using the median of each terminal node.

New in version 0.18: Mean Absolute Error (MAE) criterion.

splitter : *{"best", "random"}, default="best"*
The strategy used to choose the split at each node. Supported strategies are "best" to choose the best split and "random" to choose the best random split.

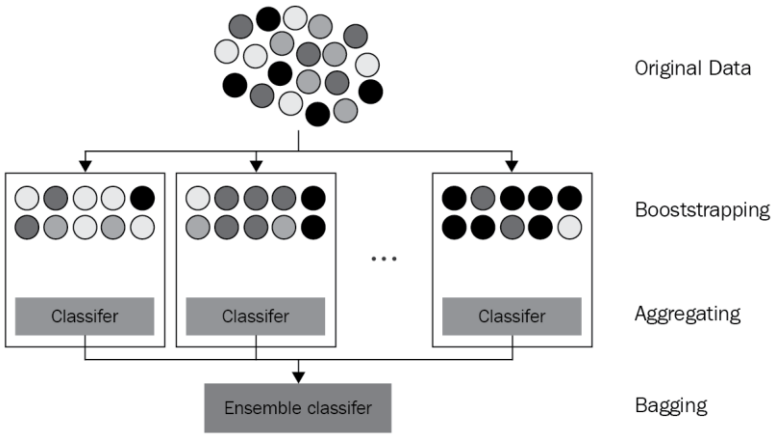
max_depth : *int, default=None*
The maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.

min_samples_split : *int or float, default=2*
The minimum number of samples required to split an internal node:

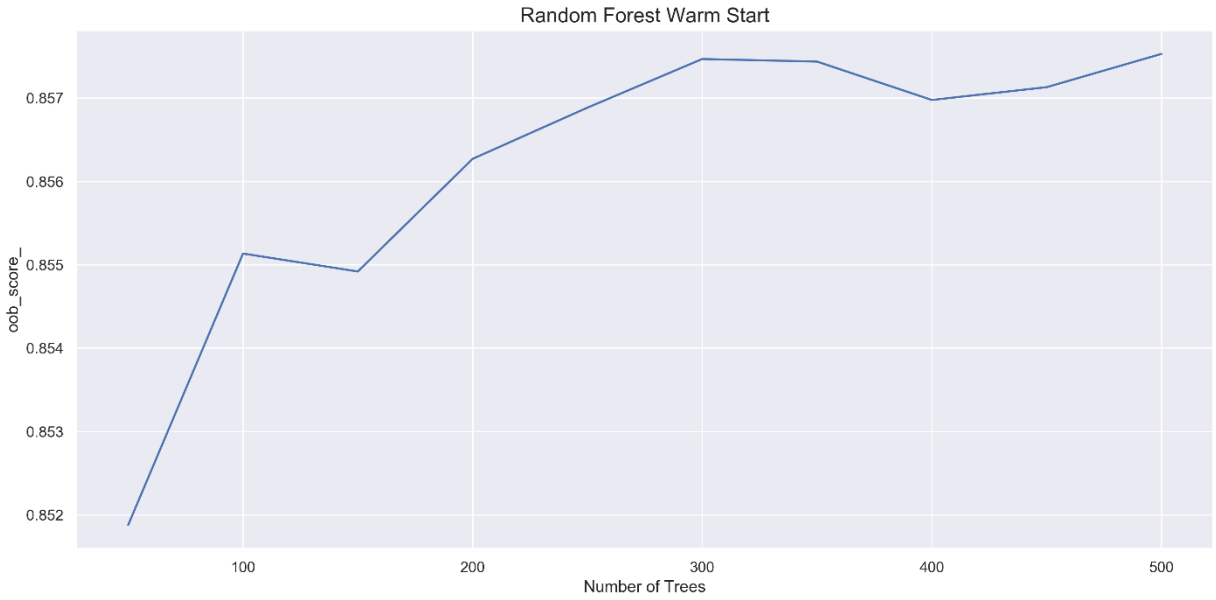
- If int, then consider `min_samples_split` as the minimum number.
- If float, then `min_samples_split` is a fraction and `ceil(min_samples_split * n_samples)` are the

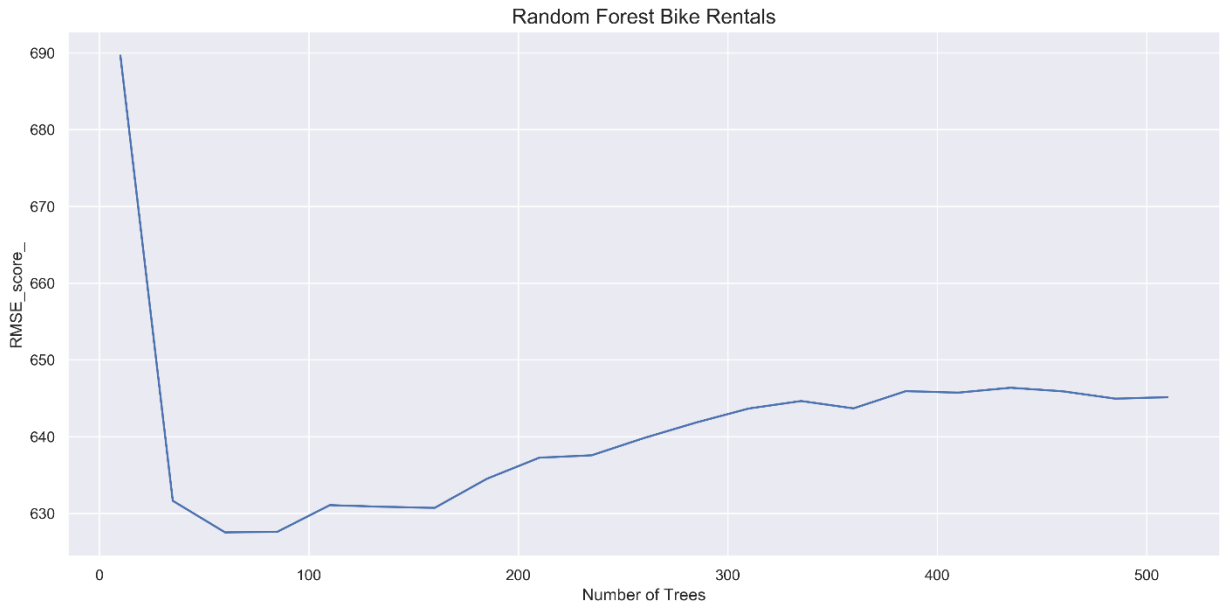
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

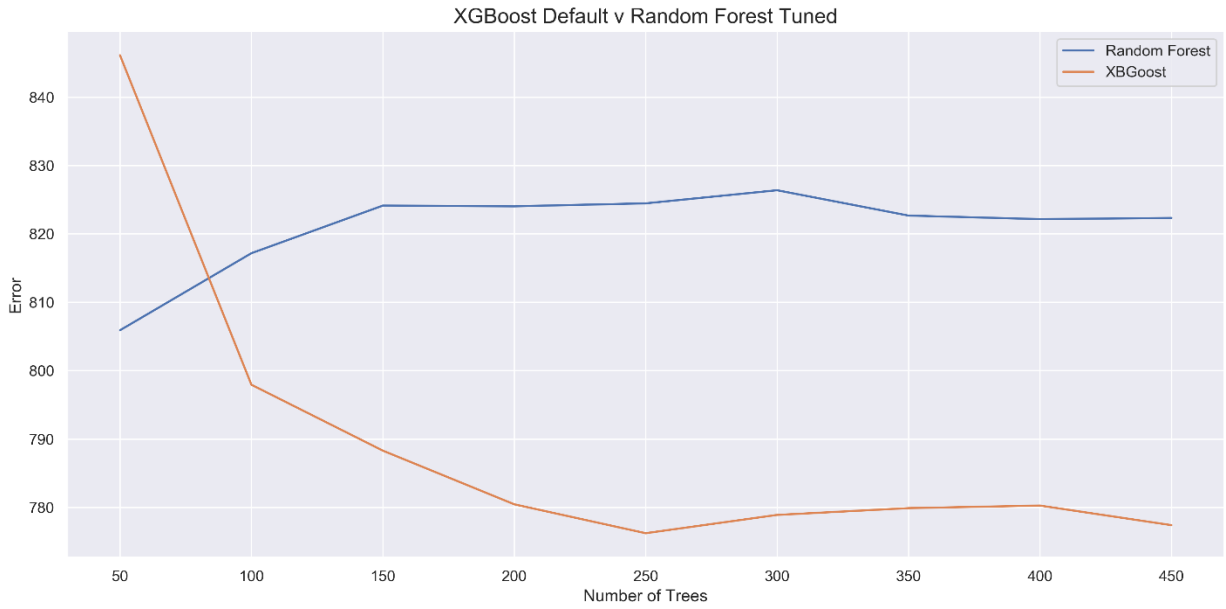
Chapter 3: Bagging with Random Forests



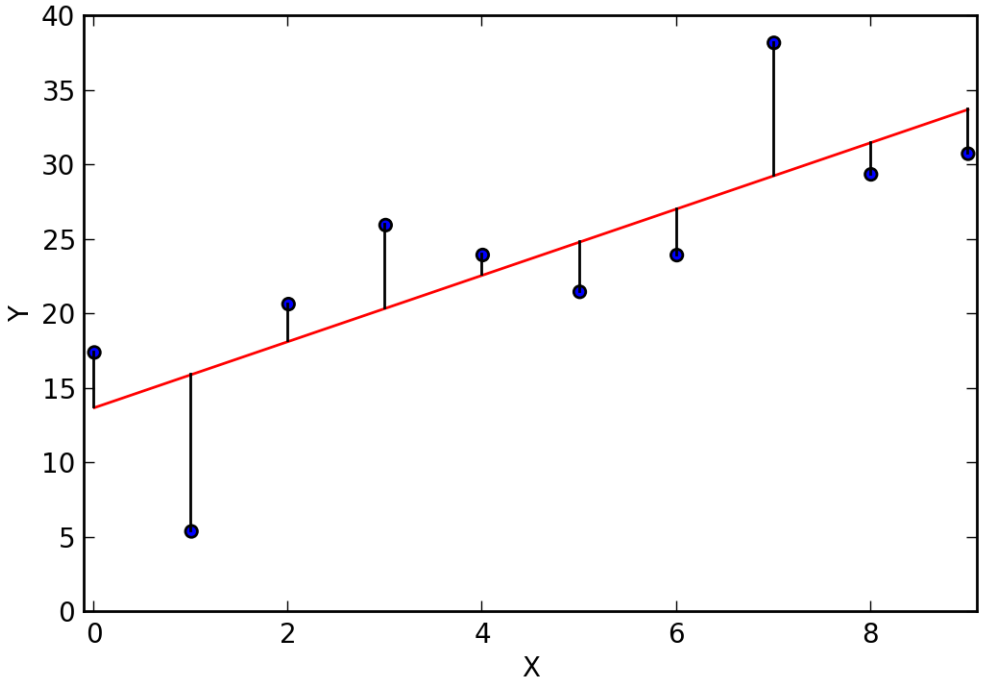
	instant	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	cnt
0	1	1.0	0.0	1.0	0.0	6.0	0.0	2	0.344167	0.363625	0.805833	0.160446	985
1	2	1.0	0.0	1.0	0.0	0.0	0.0	2	0.363478	0.353739	0.696087	0.248539	801
2	3	1.0	0.0	1.0	0.0	1.0	1.0	1	0.196364	0.189405	0.437273	0.248309	1349
3	4	1.0	0.0	1.0	0.0	2.0	1.0	1	0.200000	0.212122	0.590435	0.160296	1562
4	5	1.0	0.0	1.0	0.0	3.0	1.0	1	0.226957	0.229270	0.436957	0.186900	1600



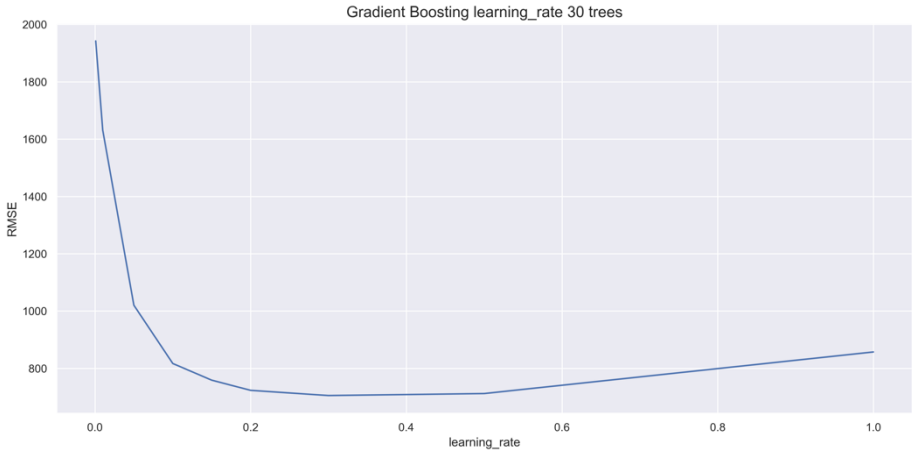

Random Forest Bike Rentals
RMSE_score_
690
680
670
660
650
640
630
0
100
200
300
400
500
Number of Trees

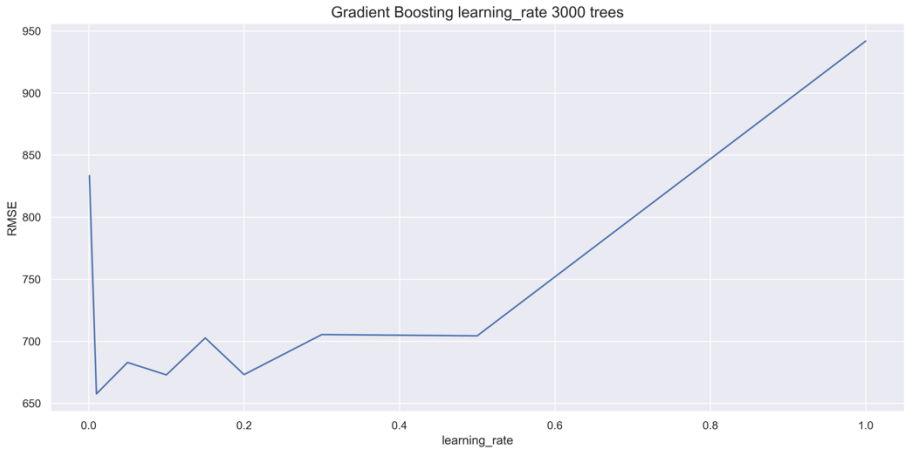

XGBoost Default v Random Forest Tuned
Random Forest
XBGoost
Error
840
830
820
810
800
790
780
50
100
150
200
250
300
350
400
450
Number of Trees

Chapter 4: From Gradient Boosting to XGBoost



	instant	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	cnt
0	1	1.0	0.0	1.0	0.0	6.0	0.0	2	0.344167	0.363625	0.805833	0.160446	985
1	2	1.0	0.0	1.0	0.0	0.0	0.0	2	0.363478	0.353739	0.696087	0.248539	801
2	3	1.0	0.0	1.0	0.0	1.0	1.0	1	0.196364	0.189405	0.437273	0.248309	1349
3	4	1.0	0.0	1.0	0.0	2.0	1.0	1	0.200000	0.212122	0.590435	0.160296	1562
4	5	1.0	0.0	1.0	0.0	3.0	1.0	1	0.226957	0.229270	0.436957	0.186900	1600





	LABEL	FLUX.1	FLUX.2	FLUX.3	FLUX.4	FLUX.5	FLUX.6	FLUX.7	FLUX.8	FLUX.9
0	2	93.85	83.81	20.10	-26.98	-39.56	-124.71	-135.18	-96.27	-79.89
1	2	-38.88	-33.83	-58.54	-40.09	-79.31	-72.81	-86.55	-85.33	-83.97
2	2	532.64	535.92	513.73	496.92	456.45	466.00	464.50	486.39	436.56
3	2	326.52	347.39	302.35	298.13	317.74	312.70	322.33	311.31	312.42
4	2	-1107.21	-1112.59	-1118.95	-1095.10	-1057.55	-1034.48	-998.34	-1022.71	-989.57

5 rows × 3198 columns

Chapter 5: XGBoost Unveiled

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	0.0
1	4.9	3.0	1.4	0.2	0.0
2	4.7	3.2	1.3	0.2	0.0
3	4.6	3.1	1.5	0.2	0.0
4	5.0	3.6	1.4	0.2	0.0

	0
count	442.000000
mean	152.133484
std	77.093005
min	25.000000
25%	87.000000
50%	140.500000
75%	211.500000
max	346.000000

PRI_jet_leading_phi	PRI_jet_subleading_pt	PRI_jet_subleading_eta	PRI_jet_subleading_phi	PRI_jet_all_pt	Weight	Label	KaggleSet	KaggleWeight
0.444	46.062	1.24	-2.475	113.497	0.000814	s	t	0.002653
1.158	-999.000	-999.00	-999.000	46.226	0.681042	b	t	2.233584
-2.028	-999.000	-999.00	-999.000	44.251	0.715742	b	t	2.347389
-999.000	-999.000	-999.00	-999.000	-0.000	1.660654	b	t	5.446378
-999.000	-999.000	-999.00	-999.000	0.000	1.904263	b	t	6.245333

	EventId	DER_mass_MMC	DER_mass_transverse_met_lep	DER_mass_vis	DER_pt_h	DER_deltaeta_jet_jet	DER_mass_jet_jet	DER_prodeta_jet_jet	DER_deltar
0	100000	138.470	51.655	97.827	27.980	0.91	124.711	2.666	
1	100001	160.937	68.768	103.235	48.146	-999.00	-999.000	-999.000	
2	100002	-999.000	162.172	125.953	35.635	-999.00	-999.000	-999.000	
3	100003	143.905	81.417	80.943	0.414	-999.00	-999.000	-999.000	
4	100004	175.864	16.915	134.805	16.405	-999.00	-999.000	-999.000	

5 rows × 33 columns

Chapter 6: Machine Learning Landscape

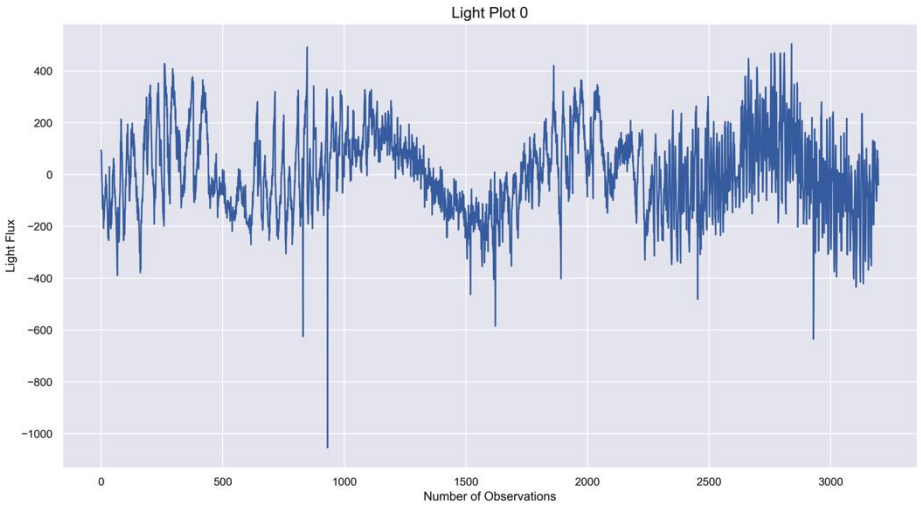
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

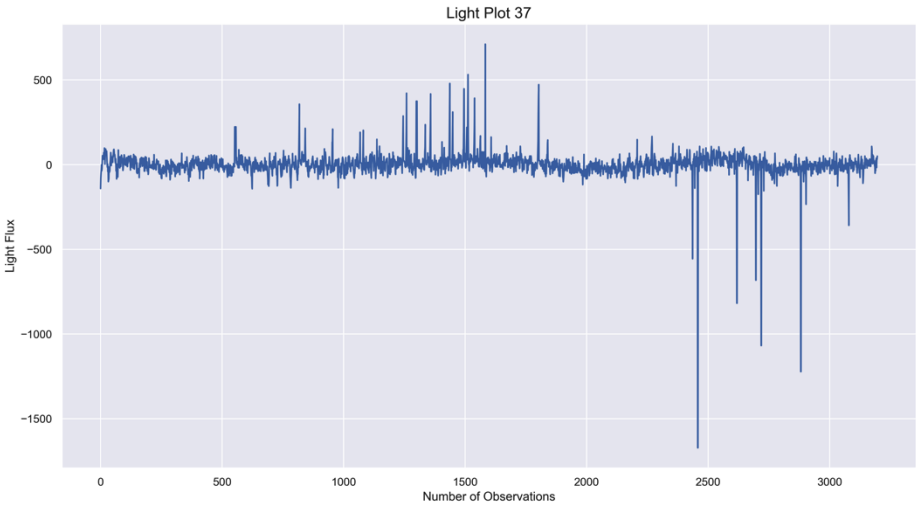
Name	Default	Range	Effect	Notes/Tips
n_estimators	100	[1, inf)	Increasing may improve scores with large data.	The number of trees in the ensemble.
learning_rate **alias:eta**	0.3	[0, inf)	Decreasing prevents overfitting.	Shrinks the tree weights in each round of boosting.
max_depth	6	[0, inf)	Decreasing prevents overfitting.	The depth of the tree. 0 is an option in a loss-guided growing policy.
gamma **alias: min_split_loss**	0	[0, inf)	Increasing prevents overfitting.	Low values, usually lower than 10, are standard.
min_child_weight	1	[0, inf)	Increasing prevents overfitting.	The minimum sum of weights required for a node to split.
subsample	1	(0, 1]	Decreasing prevents overfitting.	Limits the percentage of training rows for each boosting round.
colsample_bytree	1	(0, 1]	Decreasing prevents overfitting.	Limits the percentage of training columns for each boosting round.
colsample_bylevel	1	(0, 1]	Decreasing prevents overfitting.	Limits the percentage of columns for each depth level of the tree.
colsample_bynode	1	(0, 1]	Decreasing prevents overfitting.	Limits the percentage of columns to evaluate splits.
scale_pos_weight	1	(0, inf)	Sum(negatives)/ Sum(positives) balances data.	Used for imbalanced datasets. See *Chapter 5*, *XGBoost Unveiled*, and *Chapter 7*, *Discovering Exoplanets with XGBoost*.
max_delta_step	0	[0, inf)	Increasing prevents overfitting.	Only recommended for extremely imbalanced datasets.
lambda	1	[0, inf)	Increasing prevents overfitting.	L2 regularization of weights.
alpha	0	[0, inf)	Increasing prevents overfitting.	L1 regularization of weights.
missing	None	(-inf, inf)	Finds optimal null values.	Replace null values with numerical value like -999.0, then set equal to -999.0. See Chapter 5, XGBoost Unveiled.

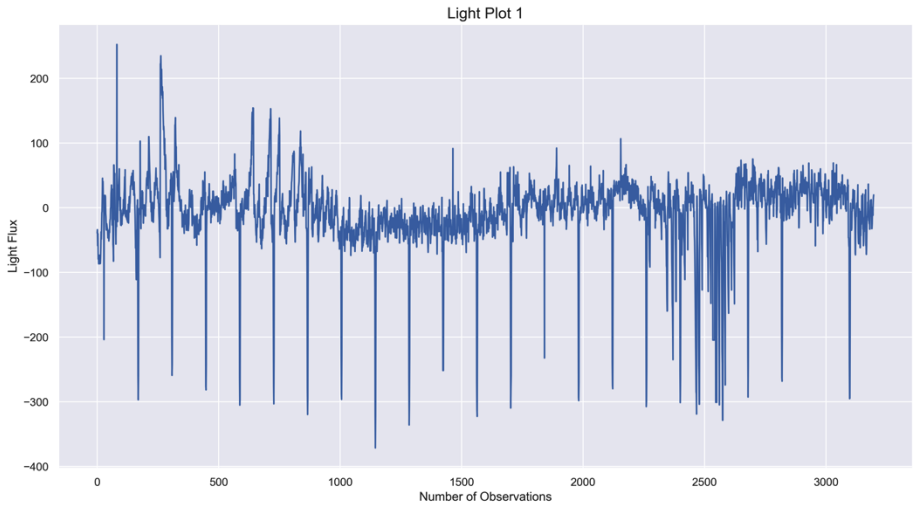
Chapter 7: Discovering Exoplanets with XGBoost

	LABEL	FLUX.1	FLUX.2	FLUX.3	FLUX.4	FLUX.5	FLUX.6	FLUX.7	FLUX.8	FLUX.9	...	FLUX.3188	FLUX.3189	FLUX.3190	FLUX.3191	FLUX.3192
0	2	93.85	83.81	20.10	-26.98	-39.56	-124.71	-135.18	-96.27	-79.89	...	-78.07	-102.15	-102.15	25.13	48.57
1	2	-38.88	-33.83	-58.54	-40.09	-79.31	-72.81	-86.55	-85.33	-83.97	...	-3.28	-32.21	-32.21	-24.89	-4.86
2	2	532.64	535.92	513.73	496.92	456.45	466.00	464.50	486.39	436.56	...	-71.69	13.31	13.31	-29.89	-20.88
3	2	326.52	347.39	302.35	298.13	317.74	312.70	322.33	311.31	312.42	...	5.71	-3.73	-3.73	30.05	20.03
4	2	-1107.21	-1112.59	-1118.95	-1095.10	-1057.55	-1034.48	-998.34	-1022.71	-989.57	...	-594.37	-401.66	-401.66	-357.24	-443.76

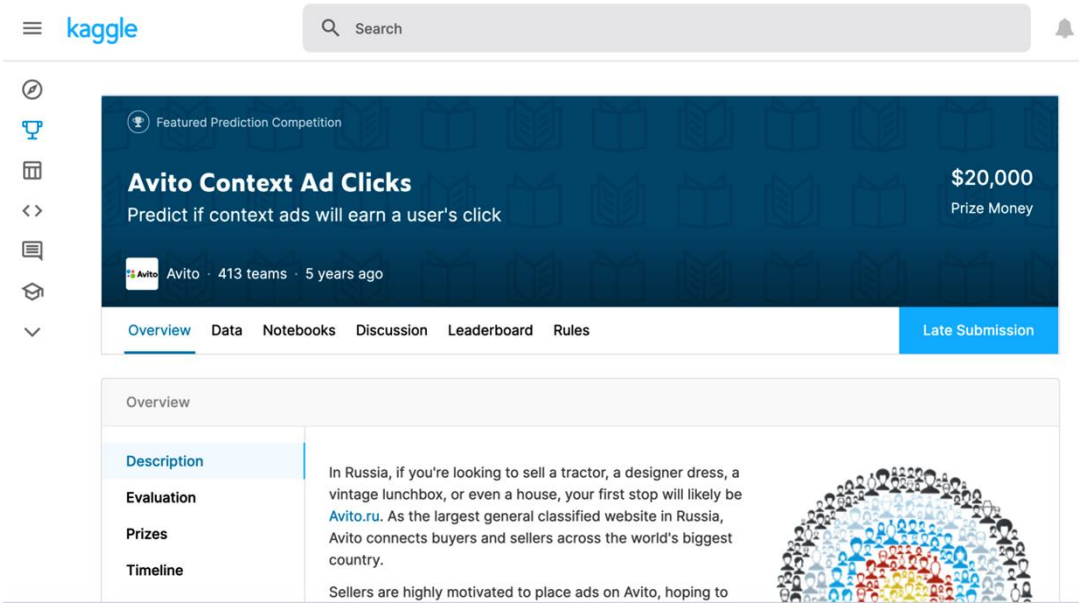
5 rows × 3198 columns




Light Plot 37
Light Flux
500
0
−500
−1000
−1500
0
500
1000
1500
2000
2500
3000
Number of Observations


Light Plot 1
Light Flux
200
100
0
−100
−200
−300
−400
0
500
1000
1500
2000
2500
3000
Number of Observations

Chapter 9: XGBoost Kaggle Masters



	distance	cab_type	time_stamp	destination	source	price	surge_multiplier	id	product_id	name
0	0.44	Lyft	1544952607890	North Station	Haymarket Square	5.0	1.0	424553bb-7174-41ea-aeb4-fe06d4f4b9d7	lyft_line	Shared
1	0.44	Lyft	1543284023677	North Station	Haymarket Square	11.0	1.0	4bd23055-6827-41c6-b23b-3c491f24e74d	lyft_premier	Lux
2	0.44	Lyft	1543366822198	North Station	Haymarket Square	7.0	1.0	981a3613-77af-4620-a42a-0c0866077d1e	lyft	Lyft
3	0.44	Lyft	1543553582749	North Station	Haymarket Square	26.0	1.0	c2d88af2-d278-4bfd-a8d0-29ca77cc5512	lyft_luxsuv	Lux Black XL
4	0.44	Lyft	1543463360223	North Station	Haymarket Square	9.0	1.0	e0126e1f-8ca9-4f2e-82b3-50505a09db9a	lyft_plus	Lyft XL

	distance	cab_type	time_stamp	destination	source	price	surge_multiplier	id	product_id	name
18	1.11	Uber	1543673584211	West End	North End	NaN	1.0	fa5fb705-03a0-4eb9-82d9-7fe80872f754	8cf7e821-f0d3-49c6-8eba-e679c0ebcf6a	Taxi
31	2.48	Uber	1543794776318	South Station	Beacon Hill	NaN	1.0	eee70d94-6706-4b95-a8ce-0e34f0fa8f37	8cf7e821-f0d3-49c6-8eba-e679c0ebcf6a	Taxi
40	2.94	Uber	1543523885298	Fenway	North Station	NaN	1.0	7f47ff53-7cf2-4a6a-8049-83c90e042593	8cf7e821-f0d3-49c6-8eba-e679c0ebcf6a	Taxi
60	1.16	Uber	1544731816318	West End	North End	NaN	1.0	43abdbe4-ab9e-4f39-afdc-31cfa375dc25	8cf7e821-f0d3-49c6-8eba-e679c0ebcf6a	Taxi
69	2.67	Uber	1543583283653	Beacon Hill	North End	NaN	1.0	80db1c49-9d51-4575-a4f4-1ec23b4d3e31	8cf7e821-f0d3-49c6-8eba-e679c0ebcf6a	Taxi

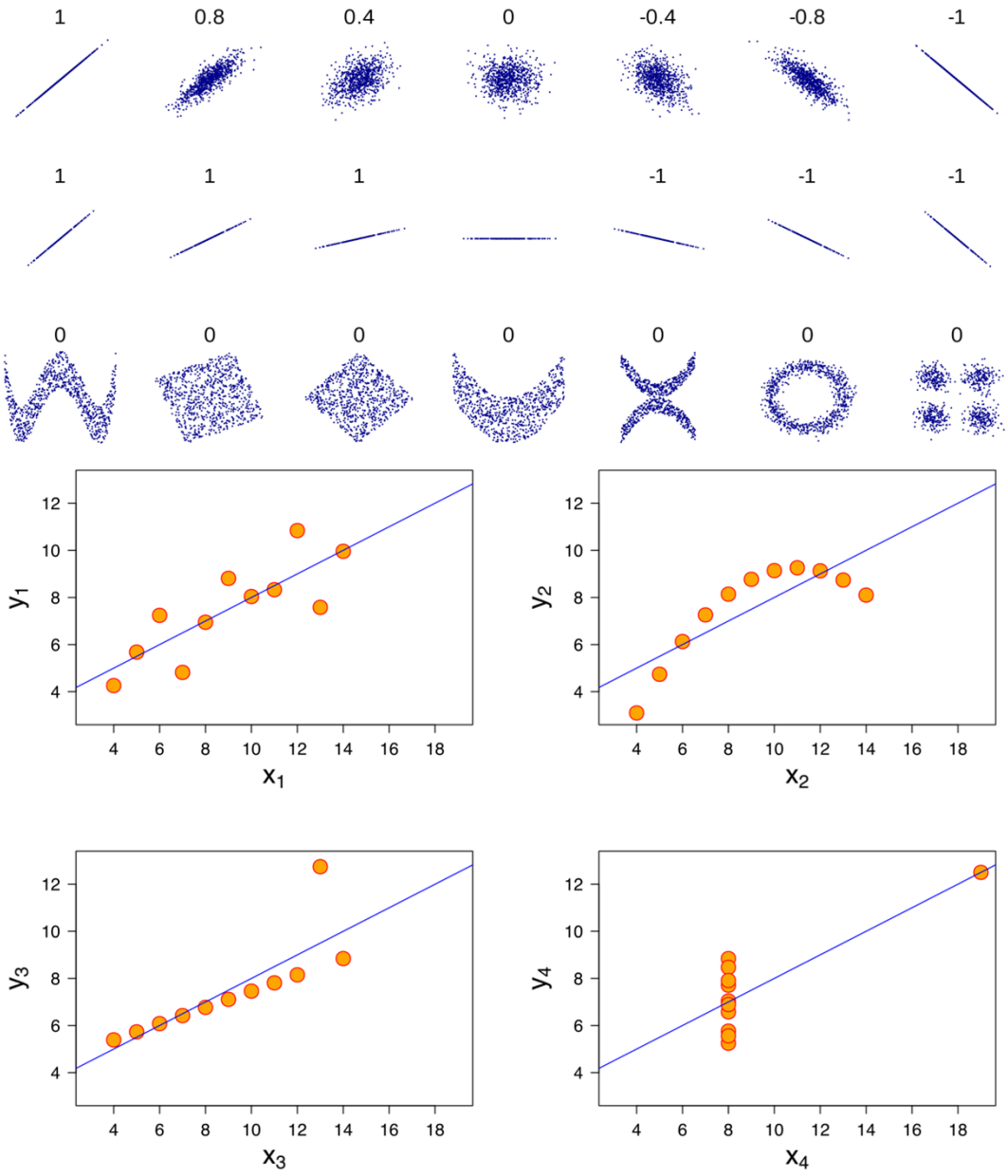
	distance	cab_type	time_stamp	destination	source	price	surge_multiplier	id	product_id	name	date
0	0.44	Lyft	1544952607890	North Station	Haymarket Square	5.0	1.0	424553bb-7174-41ea-aeb4-fe06d4f4b9d7	lyft_line	Shared	1970-01-01 00:25:44.952607890
1	0.44	Lyft	1543284023677	North Station	Haymarket Square	11.0	1.0	4bd23055-6827-41c6-b23b-3c491f24e74d	lyft_premier	Lux	1970-01-01 00:25:43.284023677
2	0.44	Lyft	1543366822198	North Station	Haymarket Square	7.0	1.0	981a3613-77af-4620-a42a-0c0866077d1e	lyft	Lyft	1970-01-01 00:25:43.366822198
3	0.44	Lyft	1543553582749	North Station	Haymarket Square	26.0	1.0	c2d88af2-d278-4bfd-a8d0-29ca77cc5512	lyft_luxsuv	Lux Black XL	1970-01-01 00:25:43.553582749
4	0.44	Lyft	1543463360223	North Station	Haymarket Square	9.0	1.0	e0126e1f-8ca9-4f2e-82b3-50505a09db9a	lyft_plus	Lyft XL	1970-01-01 00:25:43.463360223

	distance	cab_type	time_stamp	destination	source	price	surge_multiplier	id	product_id	name	date
0	0.44	Lyft	1544952607890	North Station	Haymarket Square	5.0	1.0	424553bb-7174-41ea-aeb4-fe06d4f4b9d7	lyft_line	Shared	2018-12-16 09:30:07.890
1	0.44	Lyft	1543284023677	North Station	Haymarket Square	11.0	1.0	4bd23055-6827-41c6-b23b-3c491f24e74d	lyft_premier	Lux	2018-11-27 02:00:23.677
2	0.44	Lyft	1543366822198	North Station	Haymarket Square	7.0	1.0	981a3613-77af-4620-a42a-0c0866077d1e	lyft	Lyft	2018-11-28 01:00:22.198
3	0.44	Lyft	1543553582749	North Station	Haymarket Square	26.0	1.0	c2d88af2-d278-4bfd-a8d0-29ca77cc5512	lyft_luxsuv	Lux Black XL	2018-11-30 04:53:02.749
4	0.44	Lyft	1543463360223	North Station	Haymarket Square	9.0	1.0	e0126e1f-8ca9-4f2e-82b3-50505a09db9a	lyft_plus	Lyft XL	2018-11-29 03:49:20.223

me_stamp	destination	source	price	surge_multiplier	id	product_id	name	date	month	hour	dayofweek	weekend	rush_hour
04379037	Fenway	North Station	11.5	1.0	934d2fbe-f978-4495-9786-da7b4dd21107	997acbb5-e102-41e1-b155-9df7de0a73f2	UberPool	2018-11-29 15:12:59.037	11	15	3	0	1
00477997	Fenway	North Station	26.0	1.0	af8fd57c-fe7c-4584-bd1f-beef1a53ad42	6c84fd89-3f11-4782-9b50-97c468b19529	Black	2018-12-03 01:27:57.997	12	1	0	0	0
07083241	Fenway	North Station	19.5	1.0	b3c5db97-554b-47bf-908b-3ac880e86103	6f72dfc5-27f1-42e8-84db-ccc7a75f6969	UberXL	2018-11-28 12:11:23.241	11	12	2	0	0
96813623	Fenway	North Station	36.5	1.0	fcb35184-9047-43f7-8909-f62a7b17b6cf	6d318bcc-22a3-4af6-bddd-b409bfce1546	Black SUV	2018-12-15 18:00:13.623	12	18	5	1	0
12781166	Theatre District	Northeastern University	7.0	1.0	7f0e8caf-e057-41eb-bdef-27eb14c88122	lyft_line	Shared	2018-12-03 04:53:01.166	12	4	0	0	0

destination	source	price	surge_multiplier	id	product_id	name	date	month	hour	dayofweek	weekend	rush_hour	cab_freq
Fenway	North Station	11.5	1.0	934d2fbe-f978-4495-9786-da7b4dd21107	997acbb5-e102-41e1-b155-9df7de0a73f2	UberPool	2018-11-29 15:12:59.037	11	15	3	0	1	0.504389
Fenway	North Station	26.0	1.0	af8fd57c-fe7c-4584-bd1f-beef1a53ad42	6c84fd89-3f11-4782-9b50-97c468b19529	Black	2018-12-03 01:27:57.997	12	1	0	0	0	0.504389
Fenway	North Station	19.5	1.0	b3c5db97-554b-47bf-908b-3ac880e86103	6f72dfc5-27f1-42e8-84db-ccc7a75f6969	UberXL	2018-11-28 12:11:23.241	11	12	2	0	0	0.504389
Fenway	North Station	36.5	1.0	fcb35184-9047-43f7-8909-f62a7b17b6cf	6d318bcc-22a3-4af6-bddd-b409bfce1546	Black SUV	2018-12-15 18:00:13.623	12	18	5	1	0	0.504389
Theatre District	Northeastern University	7.0	1.0	7f0e8caf-e057-41eb-bdef-27eb14c88122	lyft_line	Shared	2018-12-03 04:53:01.166	12	4	0	0	0	0.495611

source	price	surge_multiplier	id	product_id	name	date	month	hour	dayofweek	weekend	rush_hour	cab_freq	cab_type_mean
North Station	11.5	1.0	934d2fbe-f978-4495-9786-da7b4dd21107	997acbb5-e102-41e1-b155-9df7de0a73f2	UberPool	2018-11-29 15:12:59.037	11	15	3	0	1	0.504389	15.743446
North Station	26.0	1.0	af8fd57c-fe7c-4584-bd1f-beef1a53ad42	6c84fd89-3f11-4782-9b50-97c468b19529	Black	2018-12-03 01:27:57.997	12	1	0	0	0	0.504389	15.743446
North Station	19.5	1.0	b3c5db97-554b-47bf-908b-3ac880e86103	6f72dfc5-27f1-42e8-84db-ccc7a75f6969	UberXL	2018-11-28 12:11:23.241	11	12	2	0	0	0.504389	15.743446
North Station	36.5	1.0	fcb35184-9047-43f7-8909-f62a7b17b6cf	6d318bcc-22a3-4af6-bddd-b409bfce1546	Black SUV	2018-12-15 18:00:13.623	12	18	5	1	0	0.504389	15.743446
theastern Jniversity	7.0	1.0	7f0e8caf-e057-41eb-bdef-27eb14c88122	lyft_line	Shared	2018-12-03 04:53:01.166	12	4	0	0	0	0.495611	16.916357



	gbtree	dart	forest	logistic	xgb
gbtree	1.000000	0.971146	0.884584	0.914111	0.971146
dart	0.971146	1.000000	0.913438	0.914111	0.971146
forest	0.884584	0.913438	1.000000	0.943308	0.913438
logistic	0.914111	0.914111	0.943308	1.000000	0.914111
xgb	0.971146	0.971146	0.913438	0.914111	1.000000

Chapter 10: XGBoost Model Deployment

	school;sex;age;address;famsize;Pstatus;Medu;Fedu;Mjob;Fjob;reason;guardian;traveltime;studytime;failures;schoolsup;famsup;paid;activities;nursery;higher;internet;ron
0	
1	
2	
3	
4	

```
school;sex;age;address;famsize;Pstatus;Medu;Fedu;Mjob;Fjob;reason;guardian;traveltime;studytime;failures;schoolsup;famsup;paid;act
ivities;nursery;higher;internet;romantic;famrel;freetime;goout;Dalc;Walc;health;absences;G1;G2;G3
"GP";NULL;18;"U";"GT3";"A";4;4;"at_home";"teacher";"course";NULL;2;2;0;"yes";"no";"no";"no";"yes";"yes";"no";"no";4;3;4;1;1;3;4;"0
";"11";11
"GP";"F";NULL;"U";"GT3";"T";1;1;"at_home";"other";"course";"father";1;2;0;"no";"yes";"no";"no";"no";"yes";"yes";"no";5;3;3;1;1;3;2
;"9";"11";11
"GP";"F";15;"U";"LE3";"T";1;1;"at_home";"other";"other";"mother";1;2;0;"yes";"no";"no";"no";"yes";"yes";"yes";"no";4;3;2;2;3;3;6;"
12";"13";12
"GP";"F";15;"U";"GT3";"T";4;2;"health";"services";"home";"mother";1;3;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"yes";3;2;2;1;1;5;
0;"14";"14";14
"GP";"F";16;"U";"GT3";"T";3;3;"other";"other";"home";"father";1;2;0;"no";"yes";"no";"no";"yes";"yes";"no";"no";4;3;2;1;2;5;0;"11";
"13";13
"GP";"M";16;"U";"LE3";"T";4;3;"services";"other";"reputation";"mother";1;2;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"no";5;4;2;1;
2;5;6;"12";"12";13
"GP";"M";16;"U";"LE3";"T";2;2;"other";"other";"home";"mother";1;2;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";4;4;4;1;1;3;0;"13";
"12";13
"GP";"F";17;"U";"GT3";"A";4;4;"other";"teacher";"home";"mother";2;2;0;"yes";"yes";"no";"no";"yes";"yes";"no";"no";4;1;4;1;1;1;2;"1
0";"13";13
"GP";"M";15;"U";"LE3";"A";3;2;"services";"other";"home";"mother";1;2;0;"no";"yes";"no";"no";"yes";"yes";"yes";"yes";"no";4;2;2;1;1;1;0;"
15";"16";17
"GP";"M";15;"U";"GT3";"T";3;4;"other";"other";"home";"mother";1;2;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"yes";"no";5;5;1;1;1;5;0;"12
";"12";13
"GP";"F";15;"U";"GT3";"T";4;4;"teacher";"health";"reputation";"mother";1;2;0;"no";"yes";"no";"no";"yes";"yes";"yes";"yes";"no";3;3;3;1;2
;2;2;"14";"14";14
"GP";"F";15;"U";"GT3";"T";2;1;"services";"other";"reputation";"father";3;3;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"yes";"no";5;2;2;1;
1;4;0;"10";"12";13
"GP";"M";15;"U";"LE3";"T";4;4;"health";"services";"course";"father";1;1;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"yes";"no";4;3;3;1;3;5
;0;"12";"13";12
"GP";"M";15;"U";"GT3";"T";4;3;"teacher";"other";"course";"mother";2;2;0;"no";"yes";"no";"no";"yes";"yes";"yes";"yes";"no";5;4;3;1;2;3;0;
"12";"12";13
"GP";"M";15;"U";"GT3";"A";2;2;"other";"other";"home";"other";1;3;0;"no";"yes";"no";"no";"yes";"yes";"yes";"yes";"yes";4;5;2;1;1;3;0;"14"
;"14";15
"GP";"F";16;"U";"GT3";"T";4;4;"health";"other";"home";"mother";1;1;0;"no";"yes";"no";"no";"yes";"yes";"yes";"yes";"no";4;4;4;1;2;2;6;"17
";"17";17
```

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	NaN	18.0	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	4	0	11	11
1	GP	F	NaN	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	2	9	11	11
2	GP	F	15.0	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	6	12	13	12
3	GP	F	15.0	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	0	14	14	14
4	GP	F	16.0	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	0	11	13	13

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	NaN	18.0	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	4	0	11	11
1	GP	F	NaN	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	2	9	11	11

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup	famsup	paid
0	GP	NaN	18.0	U	GT3	A	4	4	at_home	teacher	course	NaN	2	2	0	yes	no	no
1	GP	F	NaN	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup	famsup	paid
0	GP	F	18.0	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no	no
1	GP	F	-999.0	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no
2	GP	F	15.0	U	LE3	T	1	1	at_home	other	other	mother	1	2	0	yes	no	no
3	GP	F	15.0	U	GT3	T	4	2	health	services	home	mother	1	3	0	no	yes	no
4	GP	F	16.0	U	GT3	T	3	3	other	other	home	father	1	2	0	no	yes	no

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0
1	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0
2	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0
3	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0
4	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0

	age	Medu	Fedu	traveltime	studytime	failures	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	18.0	4	4	2	2	0	4	3	4	1	1	3	4	0	11	11
1	-999.0	1	1	1	2	0	5	3	3	1	1	3	2	9	11	11
2	15.0	1	1	1	2	0	4	3	2	2	3	3	6	12	13	12
3	15.0	4	2	1	3	0	3	2	2	1	1	5	0	14	14	14
4	16.0	3	3	1	2	0	4	3	2	1	2	5	0	11	13	13

26	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**
0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	18.0	4.0	4.0	2.0	2.0	0.0	4.0	3.0	4.0	1.0	1.0	3.0	4.0	0.0	11.0	11.0
0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	-999.0	1.0	1.0	1.0	2.0	0.0	5.0	3.0	3.0	1.0	1.0	3.0	2.0	9.0	11.0	11.0
0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	15.0	1.0	1.0	1.0	2.0	0.0	4.0	3.0	2.0	2.0	3.0	3.0	6.0	12.0	13.0	12.0
0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	15.0	4.0	2.0	1.0	3.0	0.0	3.0	2.0	2.0	1.0	1.0	5.0	0.0	14.0	14.0	14.0
0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	16.0	3.0	3.0	1.0	2.0	0.0	4.0	3.0	2.0	1.0	2.0	5.0	0.0	11.0	13.0	13.0

	school	**sex**	**age**	**address**	**famsize**	**Pstatus**	**Medu**	**Fedu**	**Mjob**	**Fjob**	**reason**	**guardian**	**traveltime**	**studytime**	**failures**	**schoolsup**	**famsup**	**paid**
0	GP	F	18.0	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no	no
1	GP	F	-999.0	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no
2	GP	F	15.0	U	LE3	T	1	1	at_home	other	other	mother	1	2	0	yes	no	no
3	GP	F	15.0	U	GT3	T	4	2	health	services	home	mother	1	3	0	no	yes	no
4	GP	F	16.0	U	GT3	T	3	3	other	other	home	father	1	2	0	no	yes	no

26	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**
0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	18.0	4.0	4.0	2.0	2.0	0.0	4.0	3.0	4.0	1.0	1.0	3.0	4.0	0.0	11.0	11.0
0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	-999.0	1.0	1.0	1.0	2.0	0.0	5.0	3.0	3.0	1.0	1.0	3.0	2.0	9.0	11.0	11.0
0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	15.0	1.0	1.0	1.0	2.0	0.0	4.0	3.0	2.0	2.0	3.0	3.0	6.0	12.0	13.0	12.0
0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	15.0	4.0	2.0	1.0	3.0	0.0	3.0	2.0	2.0	1.0	1.0	5.0	0.0	14.0	14.0	14.0
0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	16.0	3.0	3.0	1.0	2.0	0.0	4.0	3.0	2.0	1.0	2.0	5.0	0.0	11.0	13.0	13.0