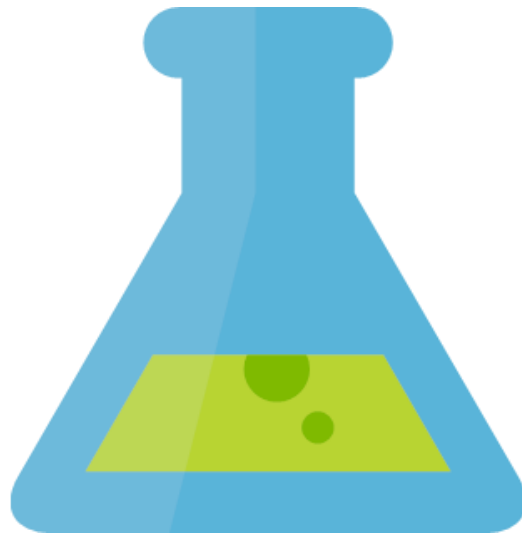


# Azure Machine Learning





# Hello!

***Il cielo è Azure sopra Berlino team***

**Fabio Rosato** - [rosato.1565173@studenti.uniroma1.it](mailto:rosato.1565173@studenti.uniroma1.it)

**Giacomo Lanciano** - [lanciano.1487019@studenti.uniroma1.it](mailto:lanciano.1487019@studenti.uniroma1.it)

**Francisco Ferreres Garcia** - [matakukos@gmail.com](mailto:matakukos@gmail.com)

**Leonardo Martini** - [martini.1722989@studenti.uniroma1.it](mailto:martini.1722989@studenti.uniroma1.it)

**Simone Caldaro** - [caldaro.1324152@studenti.uniroma1.it](mailto:caldaro.1324152@studenti.uniroma1.it)

**Na Zhu** - [nana.zhu@hotmail.com](mailto:nana.zhu@hotmail.com)

Università degli Studi di Roma “La Sapienza”

MoS in Engineering in Computer Science

Data Mining course

A.Y. 2016/2017

A decorative network diagram in the top-left corner, consisting of various sized circles (nodes) connected by thin lines (edges). Some nodes are solid grey, while others are hollow with a grey outline. The connections form a complex, branching structure.

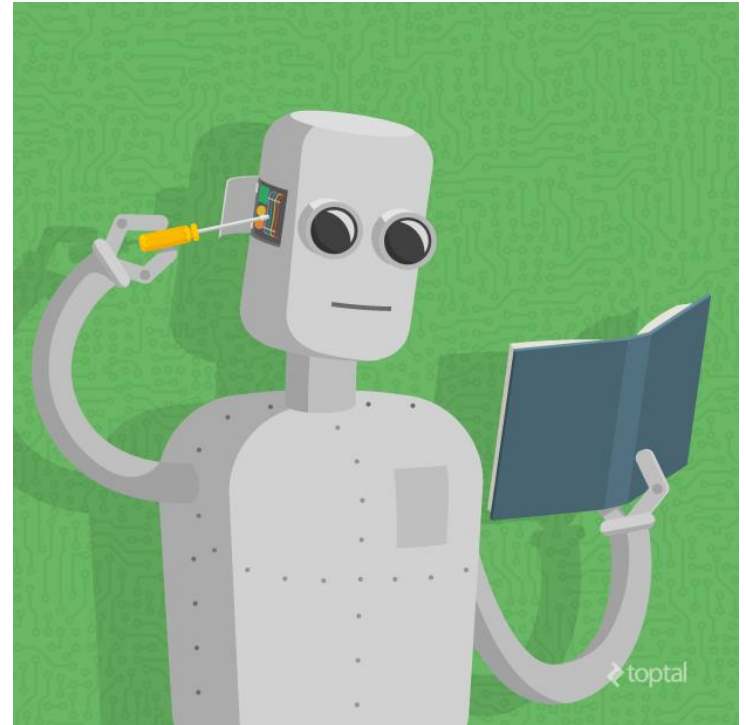
1.

# Machine Learning

A brief overview of what (the hell)  
ML means

# Machine Learning

- ◎ What is Big Data?
- ◎ What is Machine Learning?
- ◎ Uses of Machine Learning?
- ◎ Why Machine Learning?
- ◎ Who uses it?





# What is Big Data?

Structured

Unstructured



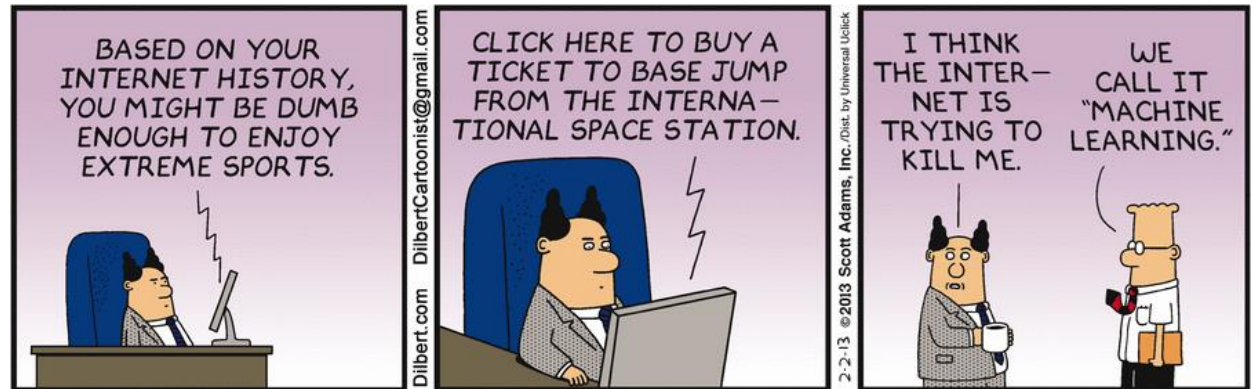
- Structured
- Unstructured

- Commercial transactions
- Social media
- Publicly available sources
- Sensors
- Business statistics

## 5

# What is machine learning?

- ◎ Examine LARGE amounts of data
  - Find patterns. Build models.
- ◎ Automatic improvement of the algorithms
  - Iterative approach.
  - Multiple passes so the machine *learns*.
- ◎ Predictions



# Uses of machine learning?

## ◎ Classification

- Supervised.
- e.g. spam filter

## ◎ Regression

- Supervised.
- Estimate relationship between continuous variables.
- e.g. car market price from specs

## ◎ Clustering

- Unsupervised.
- e.g. identify communities in social networks



# Why machine learning?

- ◎ Growing volumes and varieties of available data

- Processing this data manually would be impossible.

- ◎ Cheaper computational processing and storage

- ◎ Competitive advantage

- Companies get huge benefits by analyzing data from the markets.





# Who uses it?

- ◎ Financial institutions
  - e.g. recognize and prevent frauds.
- ◎ Governments
  - e.g. increase efficiency and service.
- ◎ Medicine and science
  - e.g. dna sequencing, patients wearable sensors.
- ◎ Marketing and sales
  - e.g. dna sequencing, patients wearable sensors.
- ◎ *You name it!*



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting a hierarchical or multi-layered structure. The lines are thin and gray, connecting the nodes in a non-linear fashion.

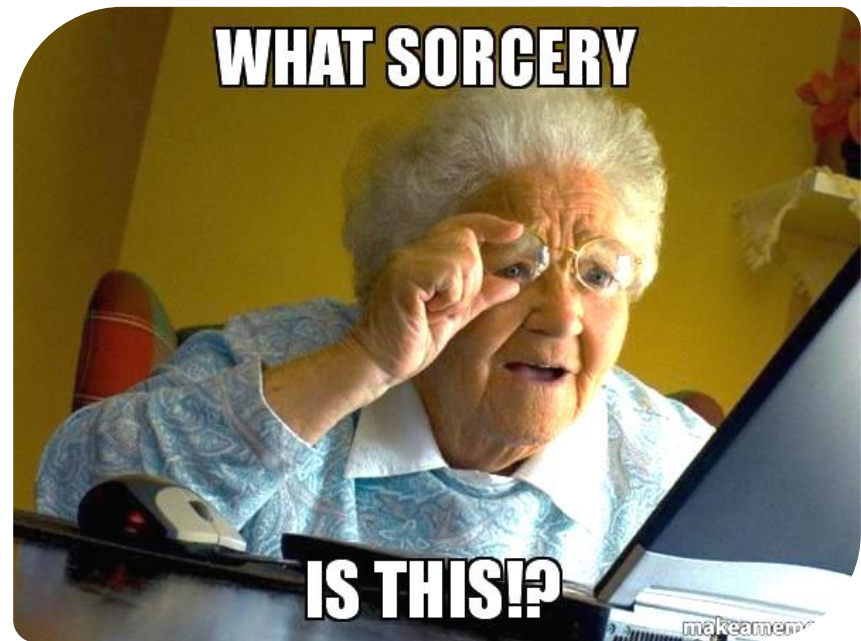
# 2.

## Using ML

A brief overview of the current tools to harness the power of ML

# ML is an incredibly powerful set of...

- ⦿ Algorithms
- ⦿ Tools
- ⦿ Techniques
- ⦿ ...
- ⦿ Magic spells?!



A decorative network diagram at the top of the slide, featuring a series of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some solid and some dashed, connected by thin lines. A central node is highlighted with a larger, dashed circle around it, containing a blue double quote symbol.

“

*Keep it simple,  
so you'll keep doing it.*

## Back in the ol' days...

To use ML, you'd have to implement the algorithms yourself:

- ⊙ prototype in some kind of friendlier language (like Matlab/Octave);
- ⊙ then implement it in a *real language* (like C++) for speed and efficiency.

## Back in the ol' days...

In-depth knowledge of ML techniques and algorithms was ***required***.

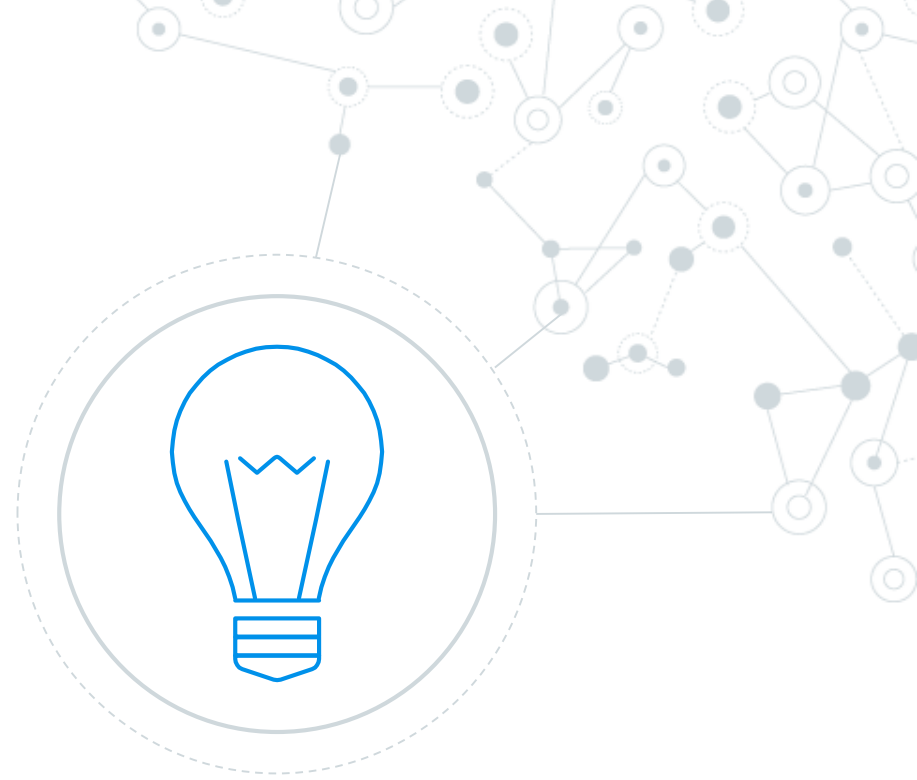
Huge barrier to adoption.



ML was used only in very big, very serious applications (that could afford and justify the *overhead*).

# Tools to the rescue!

Libraries and  
frameworks lower the  
initial effort required to  
get a working  
prototype.



# ML libraries and frameworks

- ◎ Exist for practically any widely used programming language.
- ◎ Encapsulate most widely used algorithms, abstracting away low-level details.
- ◎ Can even offer ad-hoc solutions for greater speed/efficiency/reliability (e.g. *distributed computation*).



# ML libraries and frameworks

## The celebrities:



# MLaaS

Cloud Computing  
approach gives us  
Machine Learning as a  
Service.



# ML as a Service

Outsourcing ML services:

- ◎ Incredibly low barrier to adoption.
- ◎ Massive scalability.
- ◎ *It just works!*



# ML as a Service - The celebrities:

© *Google **Prediction APIs***

© *Amazon **AWS ML***

© *Microsoft **Azure ML***

- Allows users to create and train models, then turn them into ready-to-be-consumed APIs. All through a beautifully intuitive web interface.





3.

# Azure Machine Learning Studio

Azure's solution to make your own experiments

# What is Azure Machine Learning Studio?

- ◎ Web-based workspace.
- ◎ Drag-and-drop tool.
- ◎ Collaborative environment.
- ◎ Where data science, cloud resources, and your data meet.

With Azure ML, predictive analytics solutions are...

A decorative network diagram at the top of the slide, featuring a series of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some solid and some dashed, connected by thin lines. A central node is highlighted with a larger, dashed circle around it, containing a blue double quote symbol.

“

*Easy to build.*  
*Easy to deploy.*  
*Easy to share.*

# Ease of use!

ML can do amazing things... But they could be even more amazing if **accessible to all!**





# Setup

All you need is a **web browser!** Go to Azure ML [website](#) and choose:

- ◎ **Free** workspace: start using **all the features** of Studio immediately, **no credit card** required!
- ◎ **Enterprise** workspace: add extra storage and few additional web services features (\$10/month).

Then, start working on your data from **anywhere!**

# Build

Creating a predictive model with Azure ML is as easy as ...



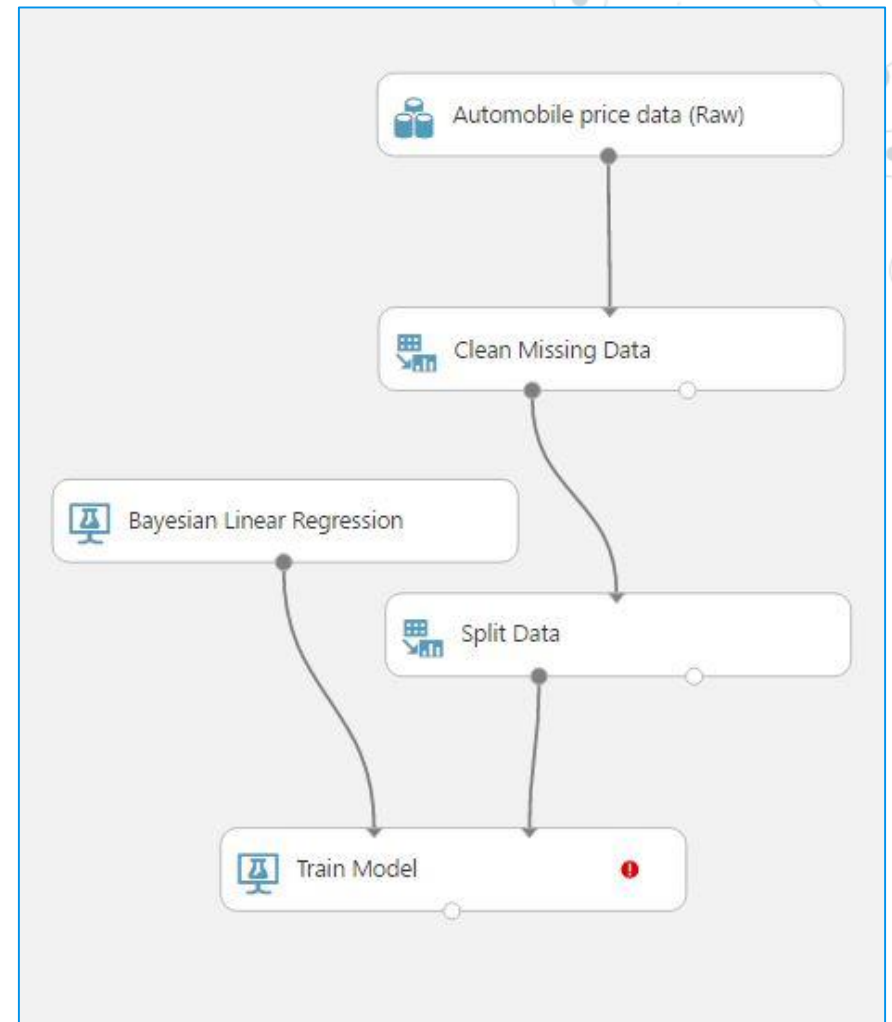
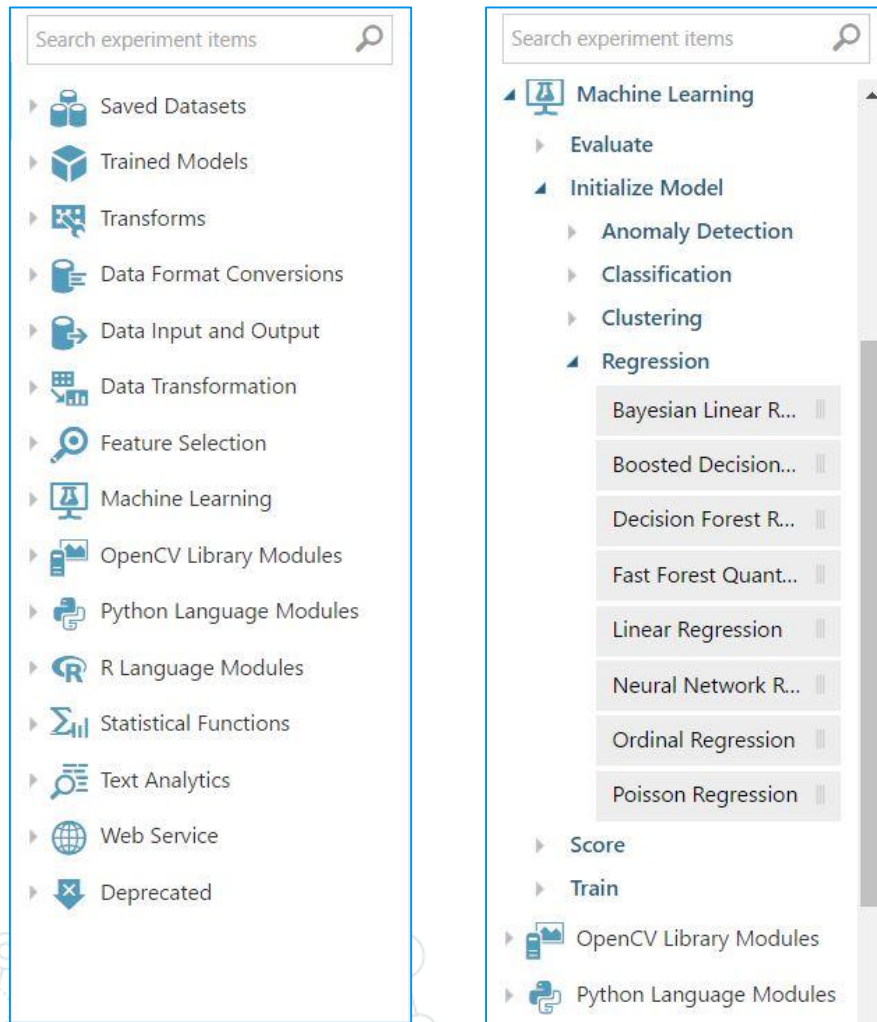
... playing with LEGO®!

## Build - main features

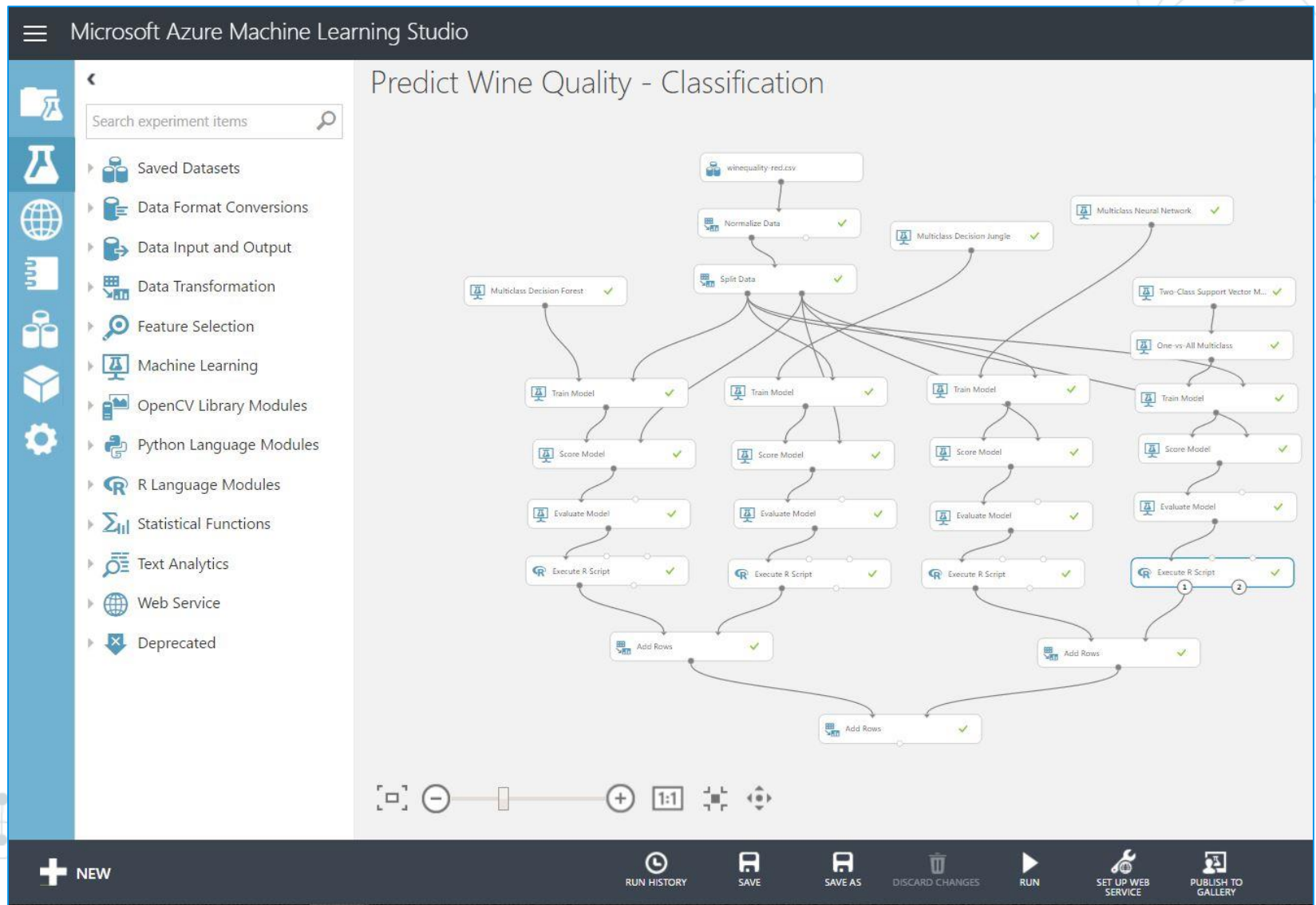
To help you building your **training experiment** (model) from scratch, Studio provides:

- ◎ Interactive, intuitive visual workspace.
- ◎ Drag-and-drop interaction to connect **modules** each other. For instance:
  - ready-to-use **datasets**.
  - ready-to-use standard **ML algorithms**.
  - your special sauce (cooked in **Python** or **R**).
  - ...
- ◎ Huge set of **samples** and **templates**.

# Build - example



# Build - advanced example



## Build - additional features


Besides creating experiments, Studio allows you to:

- ⦿ upload your own **datasets**.
- ⦿ create **web services**. (!!!)
- ⦿ store and reuse your **trained models**.
- ⦿ create Jupyter **notebooks**.
- ⦿ save your account **settings**.
- ⦿ collect all previous objects into a single **project**.



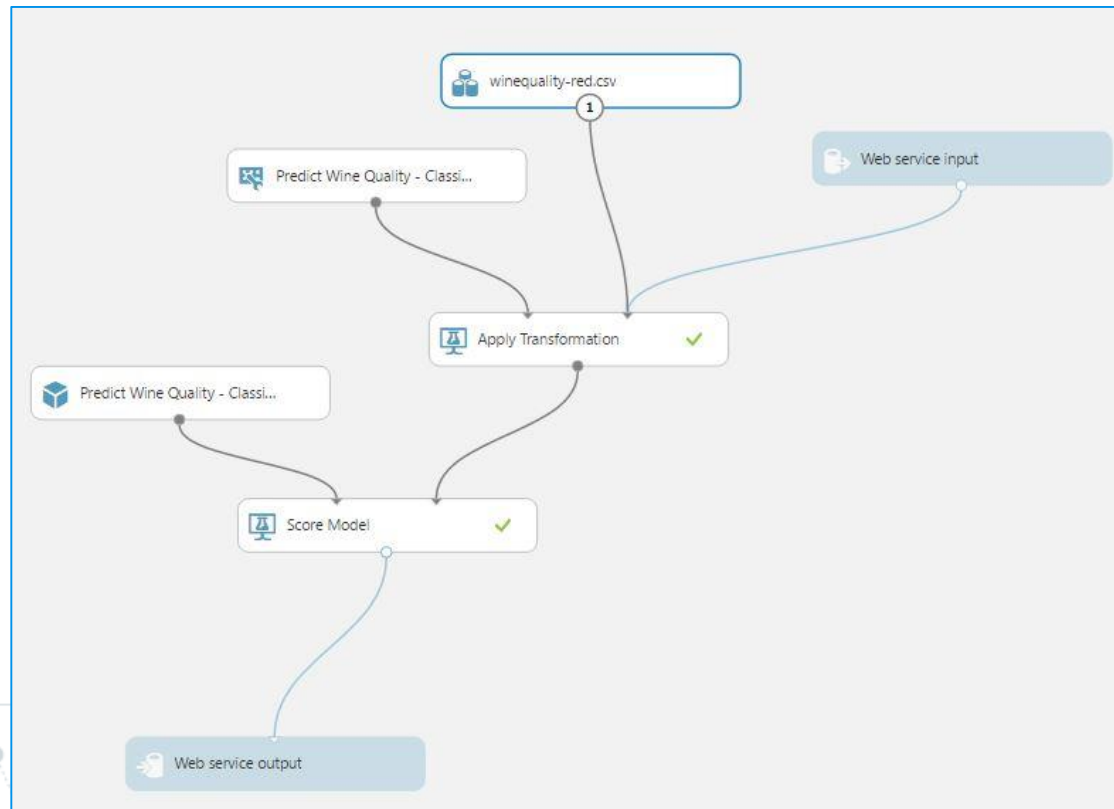
# Deploy

Once your model is ready, deploy it as a **web service** in few steps:

- ① right from Studio, click on “Setup WS”.
  - ① wait for your **predictive experiment** to be created.
  - ① click on “Deploy WS”.
  - ① wait for your **web service** to be deployed.
  - ① **enjoy!**
- 

# Deploy - predictive experiment

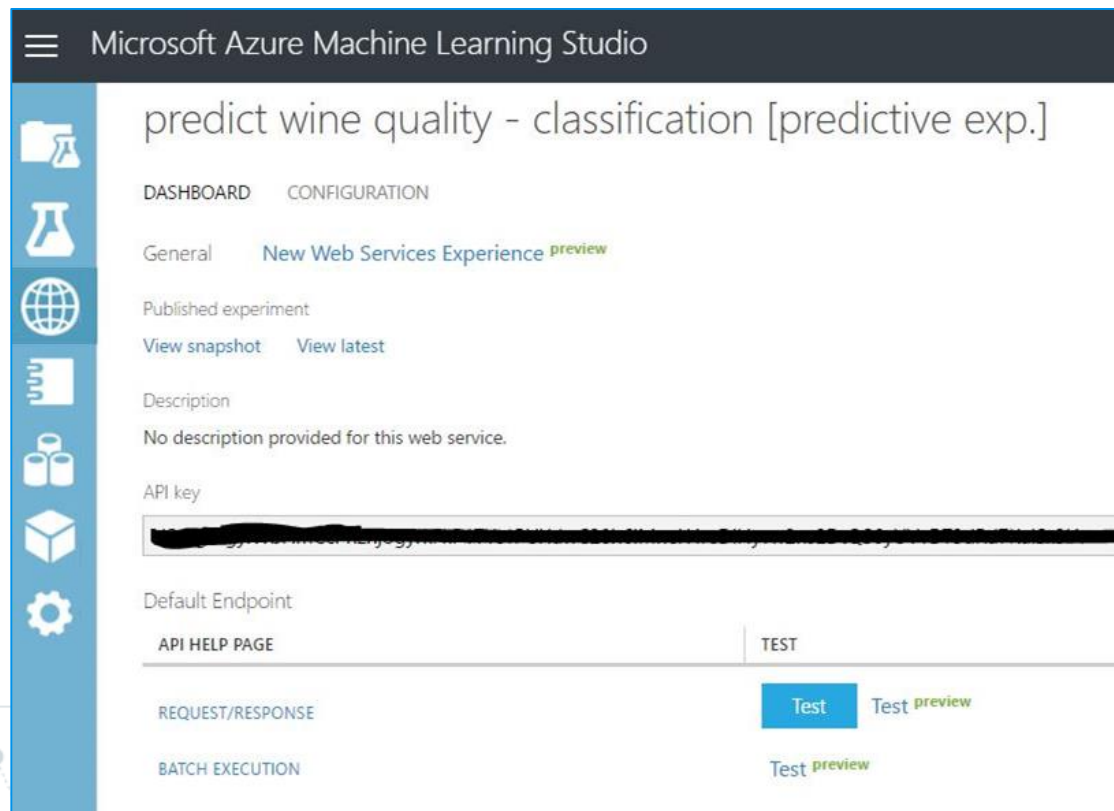
The original experiment is “translated” and the model is used to predict results.





# Deploy - web service

To call your new web service, just follow the instructions about building the **POST request**.





## Share

Your brand new experiment is ready to be shared in the community. Remember,  
**ML accessible for all!**

Upload it on **Cortana Intelligence Gallery**, where data scientists and developers share solutions.

# Share - gallery

You can publish your work directly from the Studio. Just follow the instructions and describe what you have done!

PUBLISH TO GALLERY

Experiment Description

EXPERIMENT NAME

Predict Wine Quality - Classification [Predictive Exp.]

TAG YOUR CONTENT

Press the enter key to create a tag

SUMMARY

Prediction of wine quality using Multiclass Classification analysis

DETAILED DESCRIPTION

Markdown Preview

**B** *I* | |

Enter a detailed description of your experiment

→

2 3

## Machine Learning in ML Studio

### Anomaly Detection

One-class Support Vector Machine  
Principal Component Analysis-based Anomaly Detection  
Time Series Anomaly Detection\*

### Classification

#### Two-class Classification

Averaged Perceptron  
Bayes Point Machine  
Boosted Decision Tree  
Decision Forest  
Decision Jungle  
Logistic Regression  
Neural Network  
Support Vector Machine

#### Multi-class Classification

Decision Forest  
Decision Jungle  
Logistic Regression  
Neural Network  
One-vs-all

### Clustering

K-means Clustering

### Recommendation

Matchbox Recommender

### Regression

Bayesian Linear Regression  
Boosted Decision Tree  
Decision Forest  
Fast Forest Quantile Regression  
Linear Regression  
Neural Network Regression  
Ordinal Regression  
Poisson Regression

### Statistical Functions

Descriptive Statistics  
Hypothesis Testing T-Test  
Linear Correlation  
Probability Function Evaluation

### Text Analytics

Feature Hashing  
Named Entity Recognition  
Vowpal Wabbit

### Computer Vision

OpenCV Library

<https://studio.azureml.net>

**Guest Access Workspace:** Free trial access without logging in.  
**Free Workspace:** Free persisted access, no Azure subscription needed.  
**Standard Workspace:** Full access with SLA under an Azure subscription.

Cross browser drag & drop ML workflow designer.  
Zero installation needed.

#### Data/Model Visualization

- Scatterplots
- Bar Charts
- Box plots
- Histogram
- R and Python Plotting Libraries
- REPL with Jupyter Notebook
- ROC, Precision/Recall, Lift
- Confusion Matrix
- Decision Tree\*

#### Unlimited Extensibility

- R Script Module
- Python Script Module
- Custom Module
- Jupyter Notebook

Built-in ML Algorithms

#### Training

- Cross Validation
- Retraining
- Parameter Sweep

Import Data

Preprocess

Split Data

Train Model

Score Model

Training Experiment

One-click Operationalization

Predictive Experiment

#### Make Prediction with Elastic APIs

- Request-Response Service (RRS)
- Batch Execution Service (BES)
- Retraining API

#### Data Source

- Azure Blob Storage
- Azure SQL DB
- Azure SQL DW\*
- Azure Table
- Desktop Direct Upload
- Hadoop Hive Query
- Manual Data Entry
- OData Feed
- On-prem SQL Server\*
- Web URL (HTTP)

#### Data Format

- ARFF
- CSV
- SVMlight
- TSV
- Excel
- ZIP

#### Data Preparation

- Clean Missing Data
- Clip Outliers
- Edit Metadata
- Feature Selection
- Filter
- Learning with Counts
- Normalize Data
- Partition and Sample
- Principal Component Analysis
- Quantize Data
- SQL to Transformation
- Synthetic Minority Oversampling Technique

#### Enterprise Grade Cloud Service

- SLA: 99.95% Guaranteed Up-time
- Azure AD Authentication
- Compute at Large Scale
- Multi-geo Availability
- Regulatory Compliance\*

#### Community

- Gallery (<http://gallery.azureml.net>)
- Samples & Templates
- Workspace Sharing and Collaboration
- Live Chat & MSDN Forum Support

\* Feature Coming Soon

## Azure Machine Learning Studio Capabilities Overview

© 2015 Microsoft Corporation. All rights reserved.

Created by the Azure Machine Learning Team

Email: [AzurePoster@microsoft.com](mailto:AzurePoster@microsoft.com)

Download this poster: <http://aka.ms/MLStudioOverview>



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The diagram is partially cut off by the top and left edges of the slide.

4.

# Hands-on time!

A brief tutorial about creating and deploying an experiment.

# Microsoft Azure Machine Learning Studio

- Go to Microsoft Azure Machine Learning Studio.
- In order to use the framework we need a Microsoft account:
  - A. I already have one of them  
→ just “Sign in”
  - B. I do not have any of them →  
must “Sign Up”

Welcome to Azure Machine Learning

Try it for free

No [Azure subscription](#)? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

**Sign Up** →

Already an Azure ML User?  
[Sign in here](#)

# Sign up

## Select “Free Workspace”

- ⦿ Free access
- ⦿ 10GB Storage
- ⦿ R and Python scripts support
- ⦿ Predictive web services

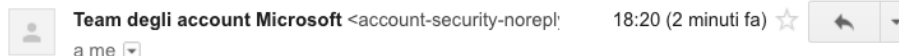
The screenshot displays the Azure ML workspace selection interface. It features three main cards: 'Quick Evaluation' (blue header), 'Most Popular' (green header), and 'Enterprise Grade' (orange header). The 'Most Popular' card is highlighted with a green border. The 'sign up here.' link in the 'Most Popular' card is circled in red. The interface includes a close button (X) in the top right corner.

Quick Evaluation	Most Popular	Enterprise Grade
Guest Workspace	Free Workspace	Standard Workspace
8-hour trial	\$0/month	\$9.99/month
No sign-in required.	Don't already have a Microsoft account? Simply <a href="#">sign up here.</a>	Azure subscription required Other charges may apply. <a href="#">Read more.</a>
<a href="#">Enter</a>	<a href="#">Sign In</a>	<a href="#">Create Workspace</a>
<ul style="list-style-type: none"><li>■ No hassle instant access</li><li>■ Stock sample datasets</li><li>■ ML models built in minutes</li><li>■ Full range of ML algorithms</li></ul>	<ul style="list-style-type: none"><li>■ Free access that never expires</li><li>■ 10 GB storage on us</li><li>■ R and Python scripts support</li><li>■ Predictive web services</li></ul>	<ul style="list-style-type: none"><li>■ Full SLA Support</li><li>■ Bring your own Azure storage</li><li>■ Parallel graph execution</li><li>■ Elastic Web Service endpoints</li></ul>



# Create an account

1. Fill the form
2. Click on create an account
3. Verify your email



Account Microsoft

## Verifica il tuo indirizzo e-mail

Per completare la configurazione di questo account Microsoft, dobbiamo verificare che questo indirizzo e-mail sia il tuo.



## Crea un account

Come nome utente per il tuo nuovo account Microsoft puoi usare qualsiasi indirizzo e-mail, tra cui gli indirizzi di Outlook.com, Yahoo! o Gmail. Se accedi già a un PC, un tablet, un telefono Windows, a Xbox Live, a Outlook.com o a OneDrive, per [accedere](#) puoi usare lo stesso account.

Nome	Cognome
<input type="text"/>	<input type="text"/>

Nome utente

<input type="text" value="Nuovo indirizzo e-mail"/>	@outlook.it ▼
---	---------------

[Usa il tuo indirizzo e-mail](#)

Password

Almeno 8 caratteri. La distinzione tra maiuscole e minuscole è rilevante

Conferma la password

Paese/area geografica

Data di nascita

Giorno ▼	Mese ▼	Anno ▼
----------	--------	--------

Sesso

Facendo clic su **Crea account** dichiari di accettare il [Contratto di Servizi Microsoft](#) e l'[informativa sulla privacy e sui cookie](#).



# Sign in

🎯 Type the account you want to use and log in in the free workspace.

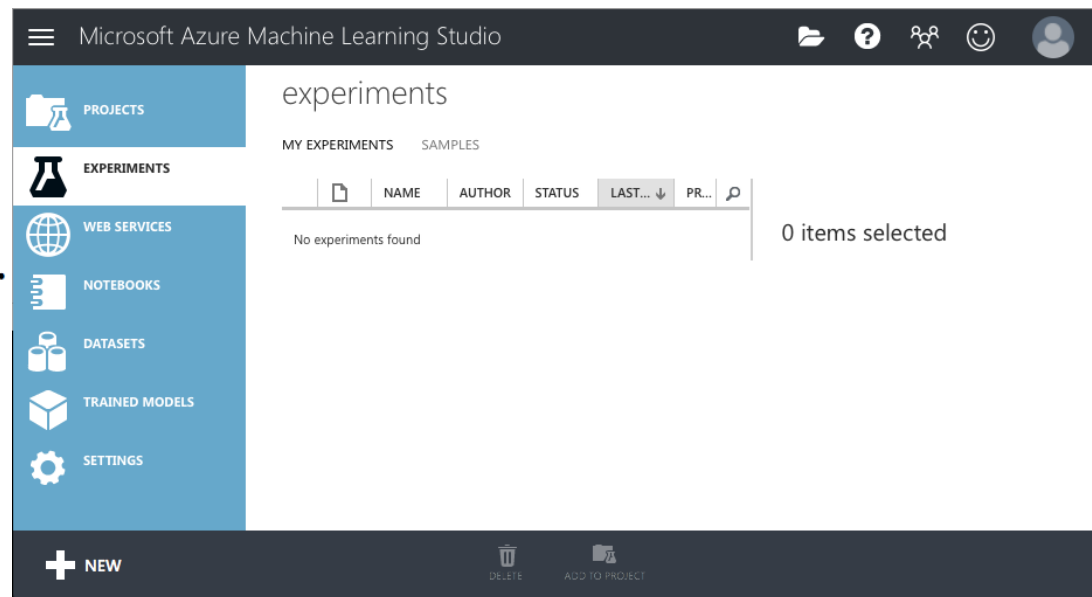
Microsoft Azure



caldaro.1324152@stu...



Use another account



# Five steps to create an experiment

## ① Create a model

- Get data
- Prepare the data
- Define features

## ② Train the model

- Choose and apply a learning algorithm

## ③ Score and test the model

- Predict new automobile prices

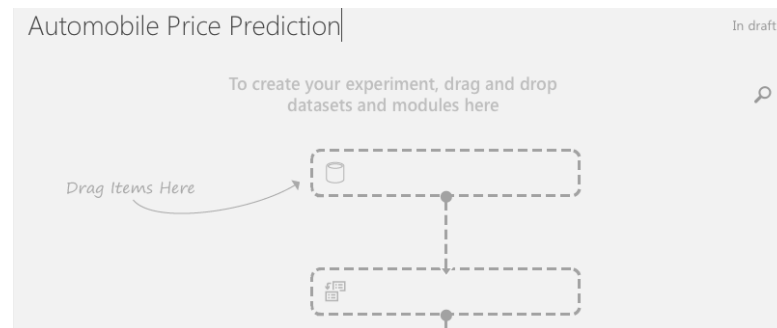
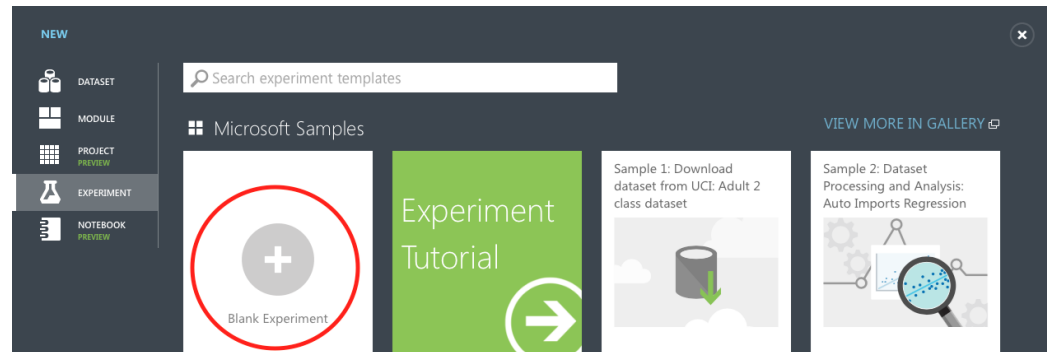
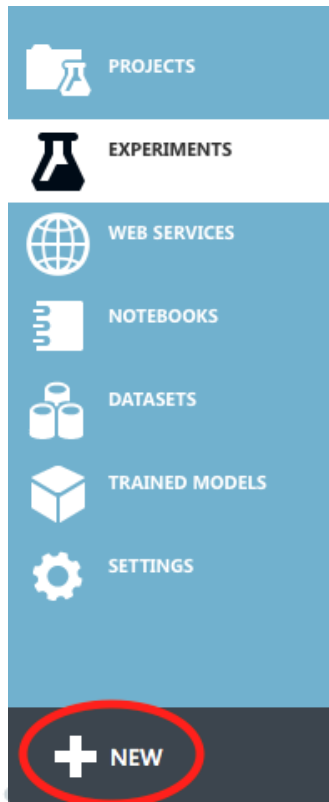
# Automobile price prediction

Technical  
Specifications



Market Price

# Create a blank experiment



# 1. Get Data

- ◎ Use data in the existing sample datasets
- ◎ Create your own dataset by NEW dataset
- ◎ Import data: Load data from sources such as the Web, Azure SQL database, Azure table, Hive table, or Windows Azure BLOB storage. Formerly known as Reader

# Using Azure saved dataset

- 🎯 In the search bar, look for automobile
- 🎯 Drag and drop the dataset in the dashboard

automobile 🔍

📦 Saved Datasets

📁 Samples

Automobile price data (...)

MPG data for various au...



Automobile Price Prediction

In draft



Automobile price data (Raw)

1

# Visualize the Data

- ☉ Selecting one column, some statistics are shown
- ☉ Given the variables for a specific automobile, we're going to try to predict the price (last column)

Automobile price data (Raw)

dataset (GenericCSV)

























Download

Visualize

Generate Data Access Code...

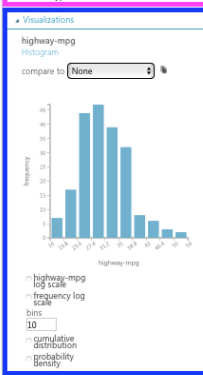
Open in a new Notebook

Automobile Price Prediction > Automobile price data (Raw) > dataset

205 columns																								
rows	26																							
make	fuel-type	aspiration	num-doors	body-style	drive-wheels	engine-location	wheel-base	length	width	height	curb-weight	engine-type	num-of-cylinders	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price	
																								
alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	13495	
alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	16500	
alfa-romero	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823	ohcv	six	152	mpfi	2.68	3.47	9	154	5000	19	26	16500	
audi	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337	ohc	four	109	mpfi	3.19	3.4	10	102	5500	24	30	13990	
audi	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2624	ohc	five	136	mpfi	3.19	3.4	8	115	5500	18	22	17450	
audi	gas	std	two	sedan	fwd	front	99.8	177.3	66.3	53.1	2507	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	15250	
audi	gas	std	four	sedan	fwd	front	105.8	192.7	71.4	55.7	2844	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	17710	
audi	gas	std	four	wagon	fwd	front	105.8	192.7	71.4	55.7	2954	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	18775	
audi	gas	turbo	four	sedan	fwd	front	105.8	192.7	71.4	55.9	3086	ohc	five	131	mpfi	3.13	3.4	8.3	140	5500	17	20	23875	
audi	gas	turbo	two	hatchback	4wd	front	99.5	178.2	67.9	52	3053	ohc	five	131	mpfi	3.13	3.4	7	160	5500	16	22	23875	
bmw	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2395	ohc	four	108	mpfi	3.5	2.8	8.8	101	5800	23	29	16430	
bmw	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2395	ohc	four	108	mpfi	3.5	2.8	8.8	101	5800	23	29	16925	
bmw	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2710	ohc	six	164	mpfi	3.31	3.19	9	121	4250	21	28	20970	
bmw	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2705	ohc	six	164	mpfi	3.31	3.19	9	121	4250	21	28	21105	
bmw	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3055	ohc	six	164	mpfi	3.31	3.19	9	121	4250	20	25	24565	
bmw	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3230	ohc	six	209	mpfi	3.62	3.39	8	182	5400	16	22	30760	
bmw	gas	std	two	sedan	rwd	front	103.5	193.8	67.9	53.7	3380	ohc	six	209	mpfi	3.62	3.39	8	182	5400	16	22	41215	
bmw	gas	std	four	sedan	rwd	front	110	197	70.9	56.3	3525	ohc	six	209	mpfi	3.62	3.39	8	182	5400	15	20	36880	
chevrolet	gas	std	two	hatchback	fwd	front	88.4	141.1	60.3	53.2	1488	l	three	61	2bbl	2.91	3.03	9.5	48	5100	47	53	5151	
chevrolet	gas	std	two	hatchback	fwd	front	94.5	155.9	63.6	52	1874	ohc	four	90	2bbl	3.03	3.11	9.6	70	5400	38	43	6295	
chevrolet	gas	std	four	sedan	fwd	front	94.5	158.8	63.6	52	1909	ohc	four	90	2bbl	3.03	3.11	9.6	70	5400	38	43	6575	
dodge	gas	std	two	hatchback	fwd	front	93.7	157.3	63.8	50.8	1876	ohc	four	90	2bbl	2.97	3.23	9.41	68	5500	37	41	5572	
dodge	gas	std	two	hatchback	fwd	front	93.7	157.3	63.8	50.8	1876	ohc	four	90	2bbl	2.97	3.23	9.4	68	5500	31	38	6377	
dodge	gas	turbo	two	hatchback	fwd	front	93.7	157.3	63.8	50.8	2128	ohc	four	98	mpfi	3.03	3.39	7.6	102	5500	24	30	7957	
dodge	gas	std	four	hatchback	fwd	front	93.7	157.3	63.8	50.6	1967	ohc	four	90	2bbl	2.97	3.23	9.4	68	5500	31	38	6229	
dodge	gas	std	four	sedan	fwd	front	93.7	157.3	63.8	50.6	1989	ohc	four	90	2bbl	2.97	3.23	9.4	68	5500	31	38	6692	
dodge	gas	turbo	sedan	fwd	front	93.7	157.3	63.8	50.6	2191	ohc	four	98	mpfi	3.03	3.39	7.6	102	5500	24	30	8558		
dodge	gas	std	four	wagon	fwd	front	103.3	174.6	64.6	59.8	2535	ohc	four	122	2bbl	3.34	3.46	8.5	88	5000	24	30	8921	
dodge	gas	turbo	two	hatchback	fwd	front	95.9	173.2	66.3	50.2	2811	ohc	four	156	mfi	3.6	3.9	7	145	5000	19	24	12964	
honda	gas	std	two	hatchback	fwd	front	86.6	144.6	63.9	50.8	1713	ohc	four	92	1bbl	2.91	3.41	9.6	58	4800	49	54	6479	
honda	gas	std	two	hatchback	fwd	front	86.6	144.6	63.9	50.8	1819	ohc	four	92	1bbl	2.91	3.41	9.2	76	6000	31	38	6855	
honda	gas	std	two	hatchback	fwd	front	93.7	150	64	52.6	1837	ohc	four	79	1bbl	2.91	3.07	10.1	60	5500	38	42	5399	
honda	gas	std	two	hatchback	fwd	front	93.7	150	64	52.6	1940	ohc	four	92	1bbl	2.91	3.41	9.2	76	6000	30	34	6529	
honda	gas	std	two	hatchback	fwd	front	93.7	150	64	52.6	1940	ohc	four	92	1bbl	2.91	3.41	9.2	76	6000	30	34	7194	

## Statistics

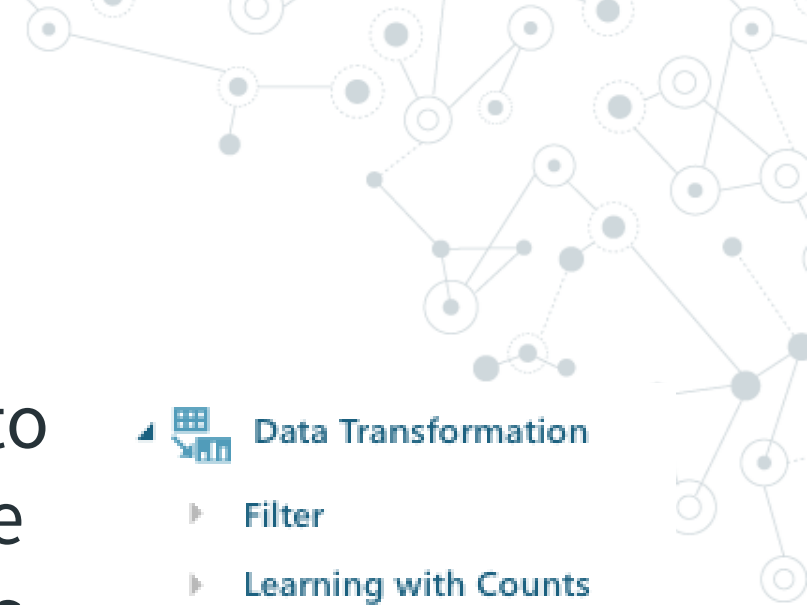


Statistics	
Mean	30.7512
Median	30
Min	16
Max	54
Standard Deviation	6.8864
Unique Values	30
Missing Values	0
Feature Type	Numeric Feature



## Visualizations

## 2. Prepare the data

🎯 This menu can be used to transform raw data to the input of the next modules

- 
- 
- ▾  Data Transformation
    - Filter
    - Learning with Counts
    - Manipulation
    - Sample and Split
    - Scale and Reduce



# Preprocess automobile dataset

1. Clean missing values present in the columns of various rows so the model can analyze the data correctly.
2. Do not consider some columns.

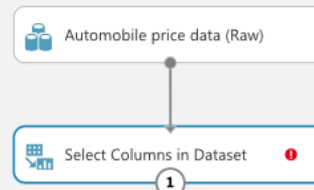
## Manipulation

Add Columns	
Add Rows	
Apply SQL Transformation	
Clean Missing Data	
Convert to Indicator Values	
Edit Metadata	
Group Categorical Values	
Join Data	
Remove Duplicate Rows	
Select Columns in Dataset	
Select Columns Transform	
SMOTE	



## Automobile Price Prediction

In draft



Properties Project

### Select Columns in Dataset

Select columns

**Selected columns:**  
Launch the selector tool to make a selection

Launch column selector

# Clean missing data: remove column

- ◎ Click on **Launch column selector**
- ◎ On the left, click **With rules**
- ◎ Under **Begin With**, click **All columns**.
- ◎ Select **Exclude** and **column names**,
- ◎ Click inside the text box and select *normalized-losses*

Select columns

BY NAME

WITH RULES

☐ Allow duplicates and preserve column order in selection

Begin With

ALL COLUMNS

NO COLUMNS

Exclude

column names

normalized-losses ✕

+ -

# Clean missing data: remove row

## Manipulation

Add Columns

Add Rows

Apply SQL Transformation

Clean Missing Data

Convert to Indicator Values

Edit Metadata

Group Categorical Values

Join Data

Remove Duplicate Rows

Select Columns in Dataset

Select Columns Transform

SMOTE

## Automobile Price Prediction



Automobile price data (Raw)



Select Columns in Dataset



Clean Missing Data

Properties Project

## Clean Missing Data

Columns to be cleaned

**Selected columns:**  
All columns

Launch column selector

Minimum missing value ra...

0

Maximum missing value r...

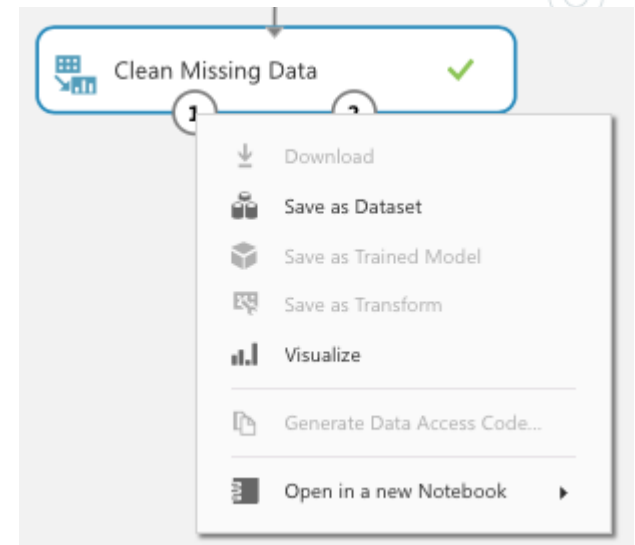
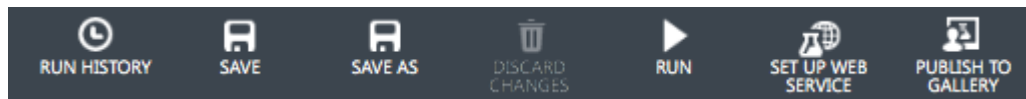
1

Cleaning mode

Remove entire row

# Run the experiment and visualize processed data

- ① Save the experiment
- ② Run it
- ③ Visualize data output from Clean Missing Data
- ④ Check differences

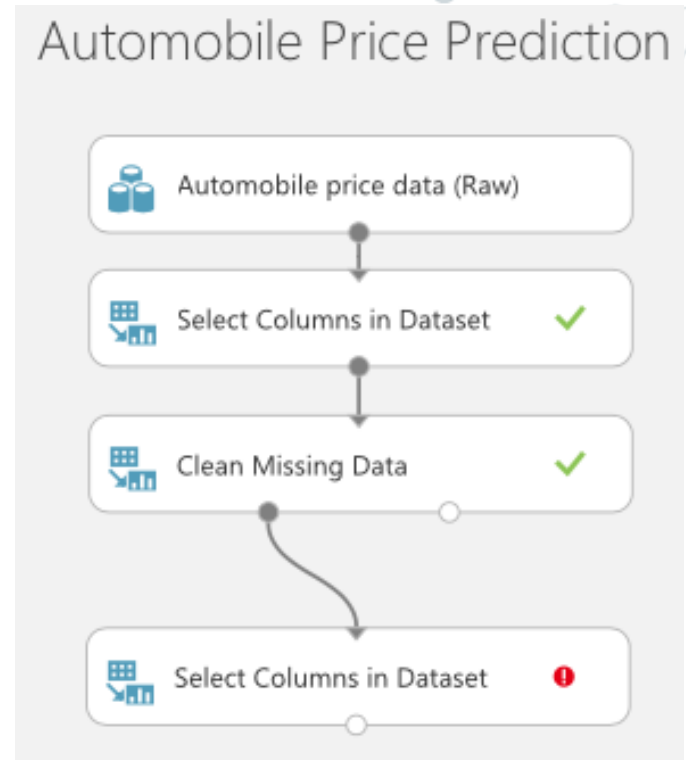


### 3. Define features

- ◎ **Features:** individual measurable properties of something you're interested in.
- ◎ Finding a good set of features for creating a predictive model requires experimentation and knowledge about the problem you want to solve.
- ◎ (In our example each row represents one automobile, and each column is a feature of that automobile)

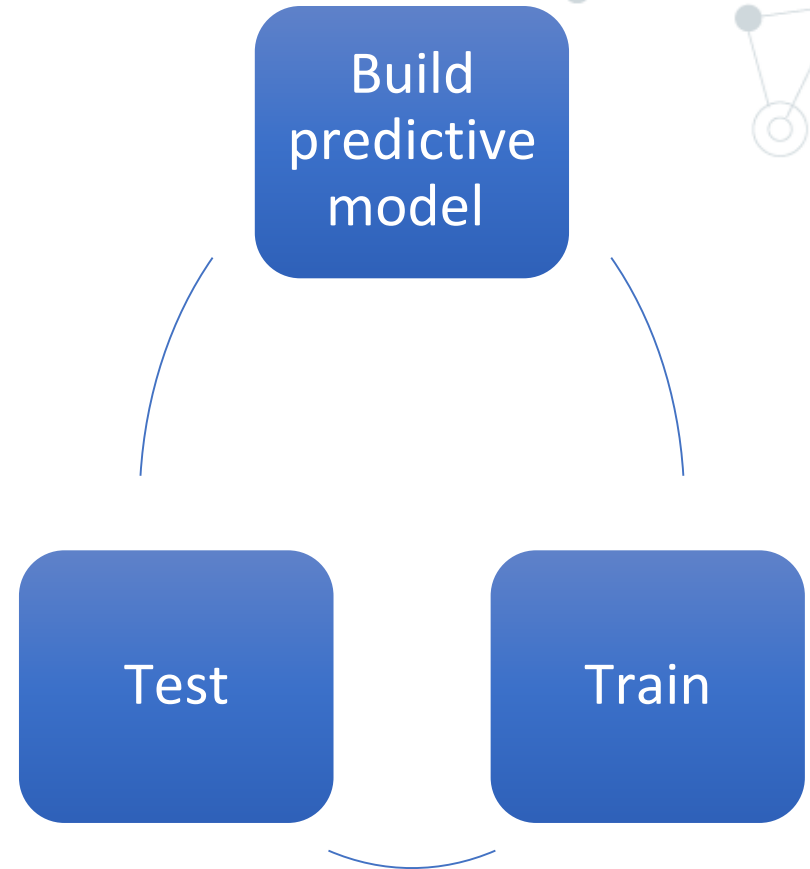
# Feature selection

- ◎ As before, drag *Select columns in Dataset*
- ◎ Connect *Clean Missing Data* to the module just added
- ◎ Click on **Launch column selector**
- ◎ On the left, click **With rules**
- ◎ Under **Begin With**, click **No columns.**
- ◎ Select **Include** and **column names**,
- ◎ Click inside the text box and select “*make*”, “*body-style*”, “*wheel-base*”, “*engine-size*”, “*horsepower*”, “*peak-rpm*”, “*highway-mpg*”, “*price*”



## 4. Choose and apply a learning algorithm

- ◎ Classification: predicts an answer from a defined set of categories
- ◎ Regression: predicts a number.
- ◎ (Because we want to predict price, which is a number, we'll use a regression algorithm)



# Split data into train set and test set

## Data Transformation

- Filter
- Learning with Counts
- Manipulation

## Sample and Split

Partition and Sample

Split Data

- Scale and Reduce



## Automobile Price Prediction



Automobile price data (Raw)



Select Columns in Dataset



Clean Missing Data



Select Columns in Dataset



Split Data



Properties Project

## Split Data

Splitting mode

Split Rows

Fraction of rows in the first...

0.75

☒ Randomized split

Random seed

0

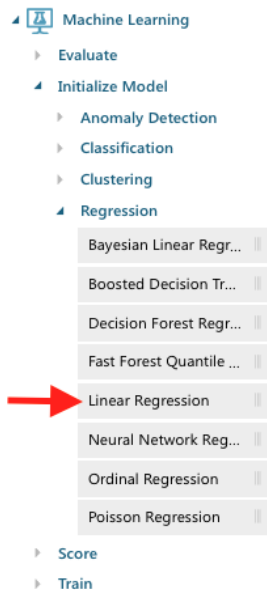
Stratified split

False

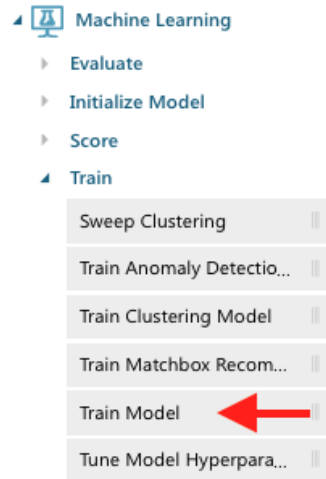


# Learning algorithm selection

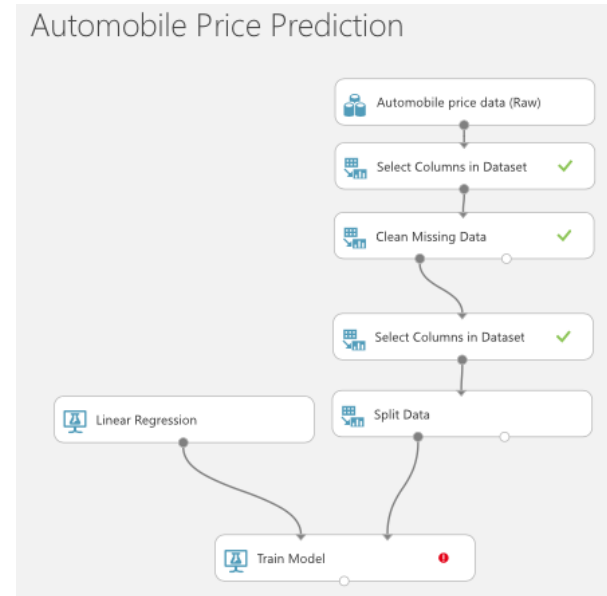
- ◎ Connect the "Train Model" module to both the "Linear Regression" and "Split Data" modules



+



=



# Train a specific feature

- Click the *Train Model* module
- Click *Launch column* selector in the Properties pane
- Click *By Name*
- Select the **price** column.
- This is the value that our model is going to predict.

Select a single column

BY NAME

WITH RULES

AVAILABLE COLUMNS

All Types search columns

make  
body-style  
wheel-base  
engine-size  
horsepower  
peak-rpm  
highway-mpg

7 columns available

SELECTED COLUMNS

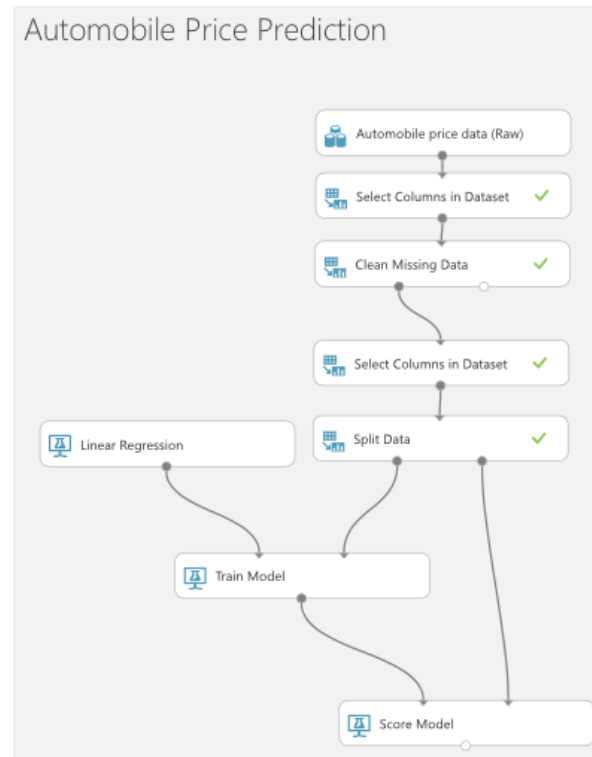
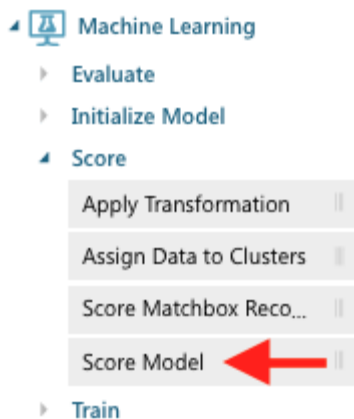
All Types search columns

price

1 columns selected

## 5. Predict new automobile prices

- 75 percent of our data used to train the model using
- 25 percent of the data to score the model functions.



# Output of the score module

## 🎯 Predicted values for price and its

Automobile Price Prediction > Score Model > Scored dataset

rows  
48

columns  
9

Real Price							Predicted Price	
make	body-style	wheel-base	engine-size	horsepower	peak-rpm	highway-mpg	price	Scored Labels
subaru	sedan	97	108	111	4800	29	11259	10286.204819
mitsubishi	hatchback	93.7	92	68	5500	38	6669	5446.847864
dodge	hatchback	93.7	90	68	5500	38	6229	6344.800711
honda	hatchback	86.6	92	76	6000	38	6855	5528.302953
alfa-romero	convertible	88.6	130	111	5000	27	16500	13498.476233
volvo	wagon	104.3	141	114	5400	28	16515	16097.608038
isuzu	hatchback	96	119	90	5000	29	11048	8315.257218
dodge	hatchback	93.7	90	68	5500	41	5572	6630.154608
bmw	sedan	101.2	108	101	5800	29	16430	19913.408695
mitsubishi	hatchback	93.7	92	68	5500	41	5389	5732.201761
bmw	sedan	103.5	209	182	5400	22	41315	30548.819502
jaguar	sedan	113	258	176	4750	19	35550	30863.486076
plymouth	hatchback	93.7	90	68	5500	38	6229	5806.676601
toyota	hatchback	102.9	171	161	5200	24	16558	17388.014192
mitsubishi	hatchback	95.9	156	145	5000	24	14489	13094.447938
plymouth	hatchback	93.7	90	68	5500	41	5572	6092.030497
volkswagen	sedan	97.3	97	52	4800	46	7995	8344.693482
dodge	hatchback	93.7	98	102	5500	30	7957	8258.383335
mercedes-benz	sedan	115.6	234	155	4750	18	34184	34960.643871
alfa-romero	hatchback	94.5	152	154	5000	26	16500	14329.816126

### Statistics

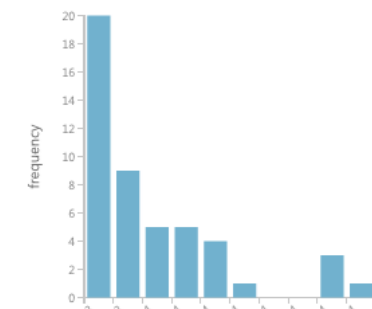
Mean	12437.776
Median	10208.7085
Min	5446.8479
Max	34960.6439
Standard Deviation	7323.458
Unique Values	46
Missing Values	0
Feature Type	Numeric Score

### Visualizations

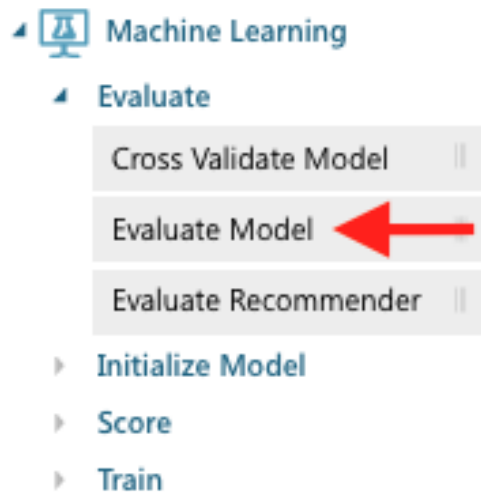
#### Scored Labels

##### Histogram

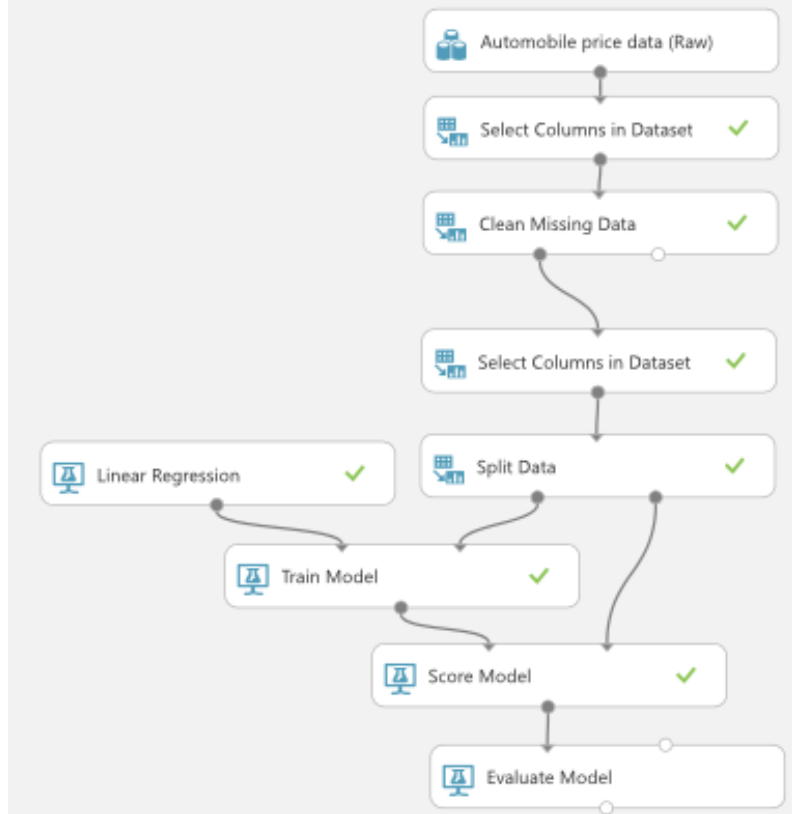
compare to



# Results evaluation



## Automobile Price Prediction



*(Final Experiment)*

# Metrics

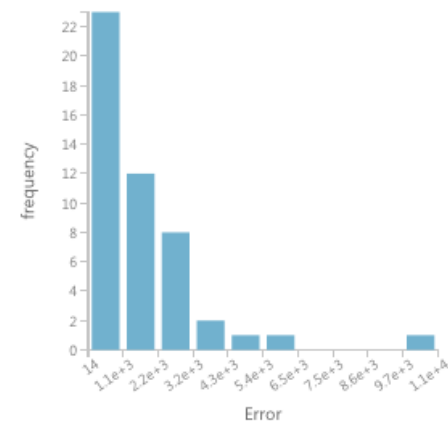
- *Mean Absolute Error (MAE)*: The average of absolute errors (an error is the difference between the predicted value and the actual value).
- *Root Mean Squared Error (RMSE)*: The square root of the average of squared errors of predictions made on the test dataset.
- *Relative Absolute Error*: The average of absolute errors relative to the absolute difference between actual values and the average of all actual values.
- *Relative Squared Error*: The average of squared errors relative to the squared difference between the actual values and the average of all actual values.
- *Coefficient of Determination*: Also known as the R squared value, this is a statistical metric indicating how well a model fits the data.

Automobile Price Prediction > Evaluate Model > Evaluation results

## Metrics

Mean Absolute Error	1656.147651
Root Mean Squared Error	2456.983209
Relative Absolute Error	0.276606
Relative Squared Error	0.089608
Coefficient of Determination	0.910392

## Error Histogram



## How a metric should be

- ◎ *For each of the error statistics, smaller is better.*
- ◎ *A smaller value indicates that the predictions more closely match the actual values.*
- ◎ *For Coefficient of Determination, the closer its value is to one (1.0), the better the predictions.*


## Iterate to improve the model

- ◎ *Change the features you use in your prediction*
- ◎ *Modify the properties of the Linear Regression algorithm*
- ◎ *Try a different algorithm altogether*
- ◎ *Add multiple machine learning algorithms to your experiment at one time*
- ◎ *Compare two of them by using the Evaluate Model module*



## 6. Deploy an Azure Machine Learning web service

- ◎ *Satisfied with your model???*
- ◎ *You can deploy it as a web service!*
- ◎ *Use the WebService to predict automobile prices by using new data...*



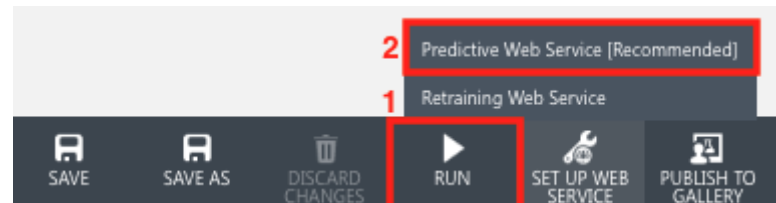
Create a training experiment

Convert the training experiment to a predictive experiment

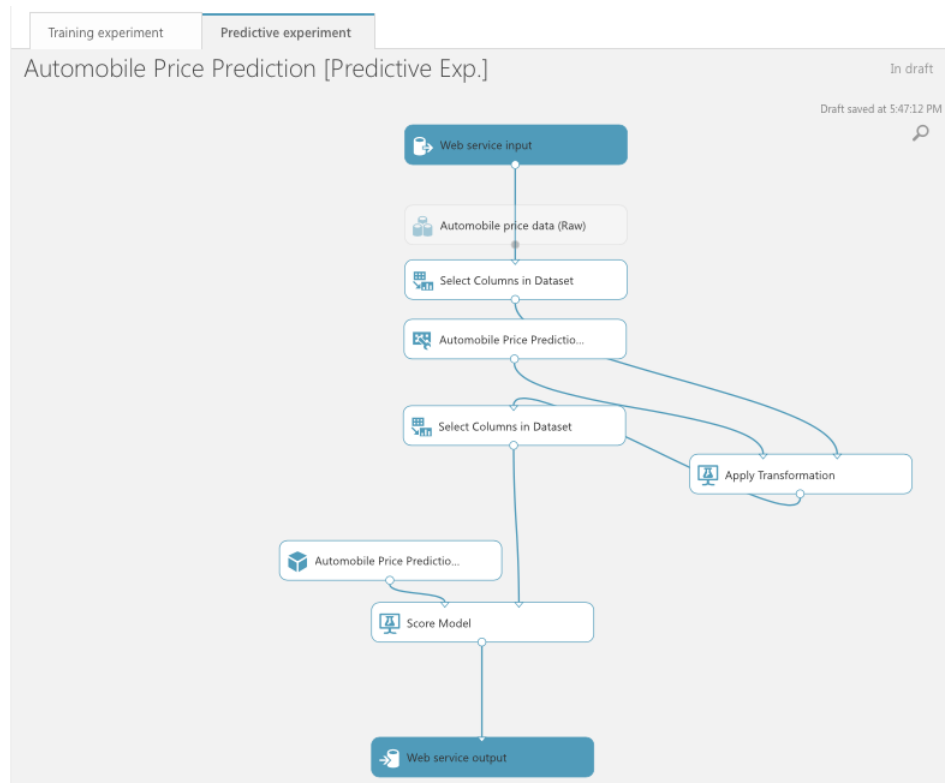
Deploy the predictive experiment as a New web service

# Convert the training experiment to a predictive experiment

- ◎ *By converting to a predictive experiment, you're getting your trained model ready to be deployed as a scoring web service.*
- ◎ *Users of the web service can send input data to your model and your model will send back the prediction results.*
- ◎ *As you convert to a predictive experiment, keep in mind how you expect your model to be used by others.*

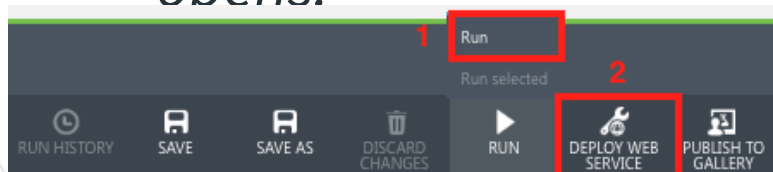


# Predictive experiment



# Deploy the predictive experiment as a New web service

- ① Click Run
- ② Click Deploy Web Service
- ③ Select Deploy Web Service New.
- ④ The deployment page of the Machine Learning Web Service portal opens.



automobile price prediction [predictive exp.]

DASHBOARD CONFIGURATION

General [New Web Services Experience preview](#)

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

vmXLRKhaeo4fxYtekazgEZ16VAKqGxP9LPp+Dkm25nK

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDAT...
REQUEST/RESPONSE	Test Test preview	Excel 2013 or later	12/9/2016 6:13:09 PM
BATCH EXECUTION	Test preview	Excel 2013 or later w	12/9/2016 6:13:09 PM

# Test your Web Service with a Python Program



web services

NAME	CREATED ON	PROJECT
Automobile Price Prediction (Predictive E...	12/9/2016 6:13:06 PM	None

- request/response page contains Request Response API Documentation, with a starter Python program (that must be modified) to call the web service

automobile price prediction [predictive exp.]

DASHBOARD CONFIGURATION

General [New Web Services Experience](#) [preview](#)

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

**Remember this... You'll need it**

API key

vmXLRKhaeo4fxYtekazgEZ16VAKqGxP9LPp+Dkm25nKY3EQMmGd9k0NWjS9mlhuKWnzz4f02YxC

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
<a href="#">REQUEST/RESPONSE</a>	<a href="#">Test</a> <a href="#">Test preview</a>	<a href="#">Excel 2013 or later</a> <a href="#">Excel 2010 or earlier</a>	12/9/2016 6:13:09 PM
<a href="#">BATCH EXECUTION</a>	<a href="#">Test preview</a>	<a href="#">Excel 2013 or later workbook</a>	12/9/2016 6:13:09 PM

# Available material



<https://github.com/giacomolanciano/Azure-Machine-Learning-tutorial>



<http://www.slideshare.net/GiacomoLanciano/azure-machine-learning-tutorial>

# Thanks!

**Any questions?**

