Azure
Machine Learning

Business Intelligence

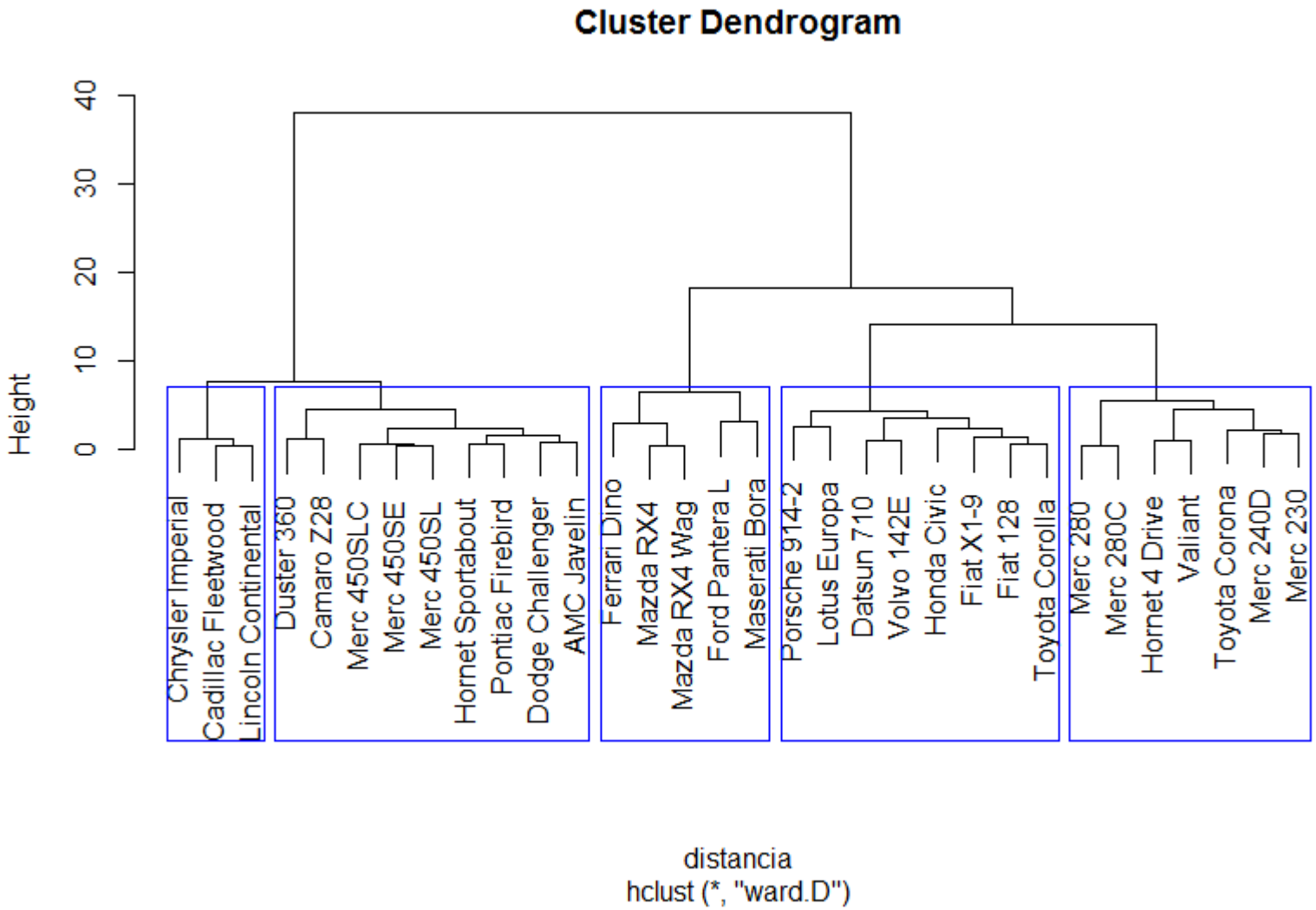
Azure Machine Learning

Prof. Leandro Guerra

E-mail: leandro.guerra@outspokenmarket.com.br **IG: @leandrowar**

Dendogramas no R

Um pouco mais...



Dendogramas no R

Um pouco mais...

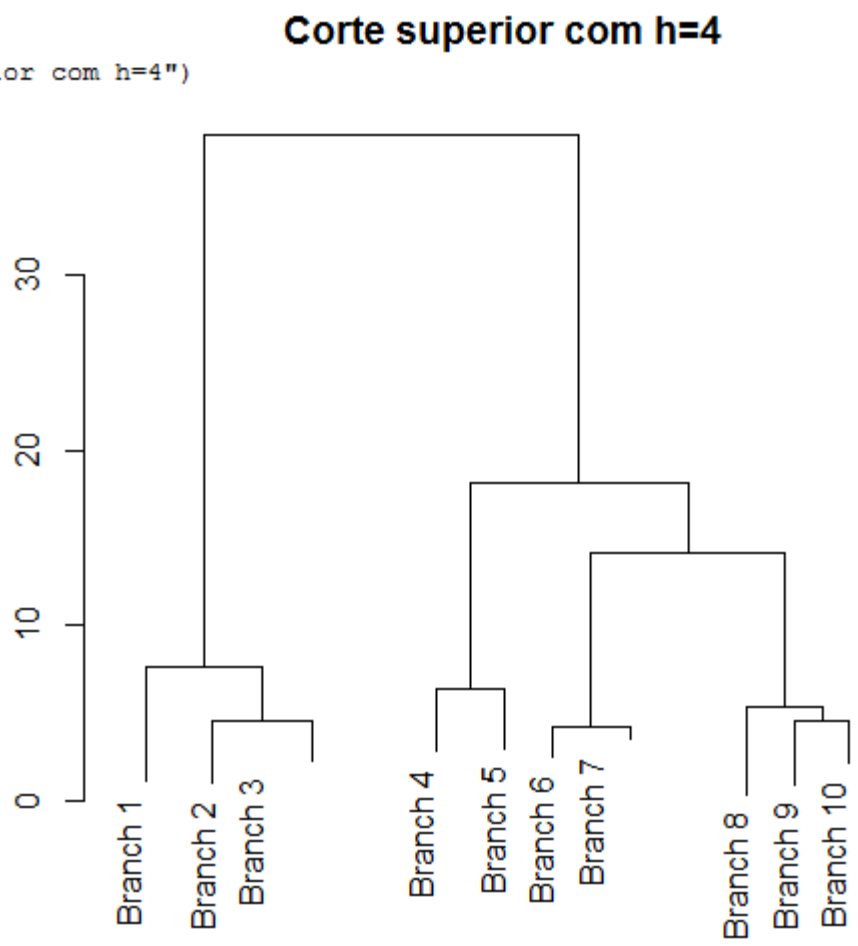
```
> grupos
          Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive   Hornet Sportabout             Valiant
                  1                   1                   2                   3                   4                   3
         Duster 360           Merc 240D            Merc 230            Merc 280           Merc 280C          Merc 450SE
                  4                   3                   3                   3                   3                   4
         Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental   Chrysler Imperial            Fiat 128
                  4                   4                   5                   5                   5                   2
        Honda Civic      Toyota Corolla       Toyota Corona    Dodge Challenger         AMC Javelin          Camaro Z28
                  2                   2                   3                   4                   4                   4
   Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa      Ford Pantera L        Ferrari Dino
                  4                   2                   2                   2                   1                   1
      Maserati Bora          Volvo 142E
                  1                   2
```

Dendogramas no R

Um pouco mais...

```
#Exibe o dendrograma com menos níveis
#Primeiro convertemos em um objeto de dendograma
dendo <- as.dendrogram(dendo)

#Agora o plot
plot(cut(dendo , h = 4)$upper, main = "Corte superior com h=4")
```



Cloud Computing

Conceito

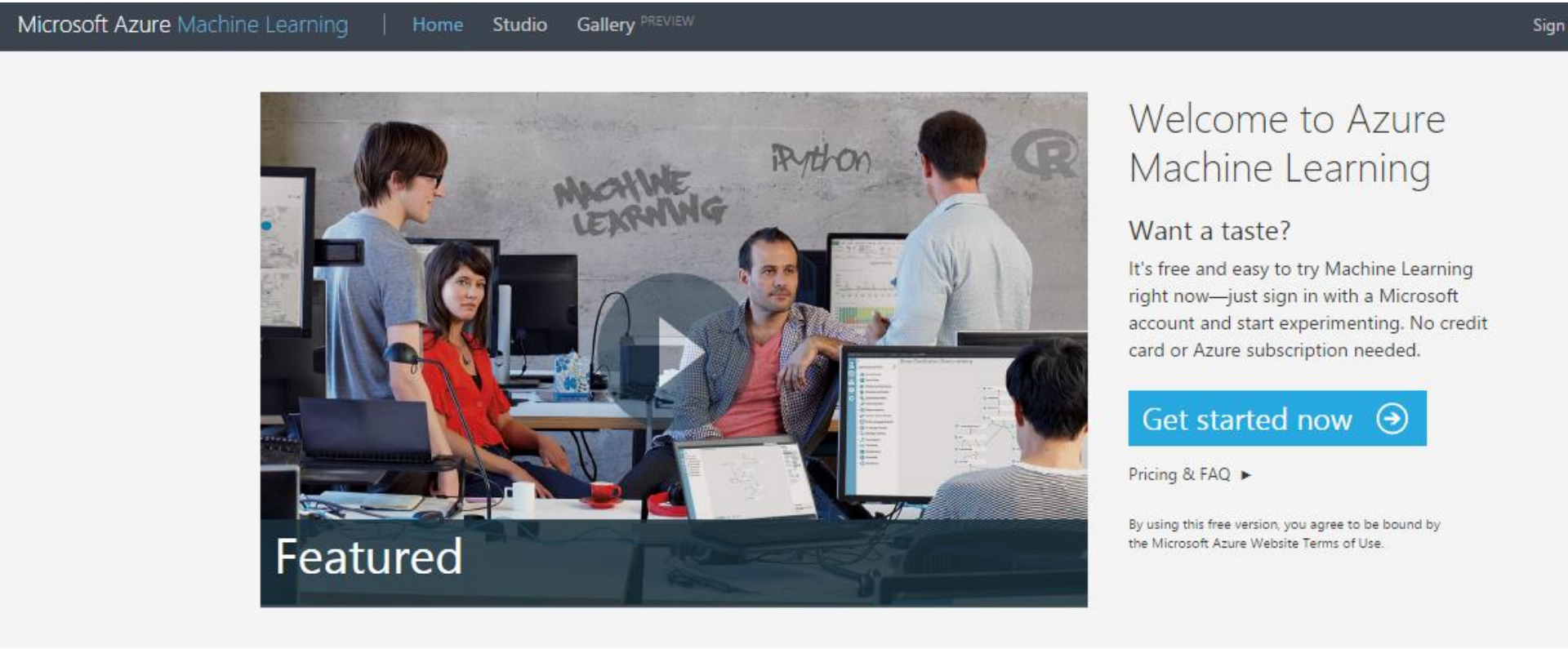
"São grupos de servidores remotos e redes de software que permitem armazenamento e processamento de dados centralizados, uqe podem ser acessados com um computador e conexão à internet. Os ambiente de Cloud computing podem ser privados, públicos ou híbridos"

**http://en.wikipedia.org/wiki/Cloud_computing*

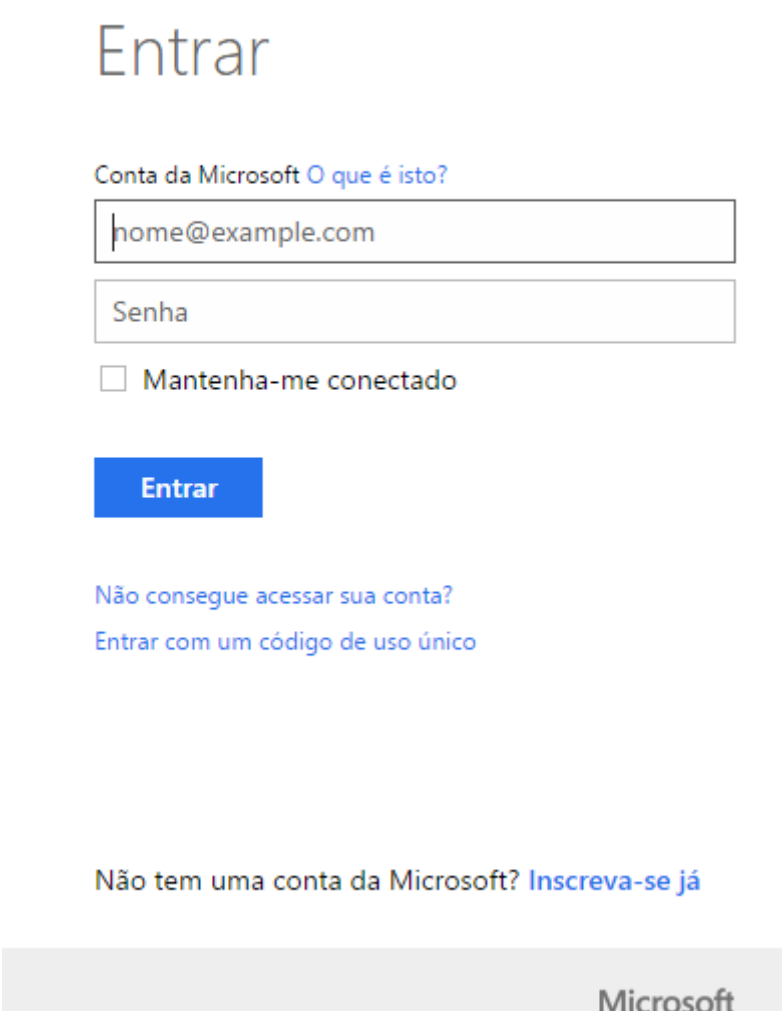
Cloud Computing
Alguns provedores

*http://www.intermedia.net/blog/2014/04/01/from-osterman-research-how-to-choose-a-cloud-communications-provider/

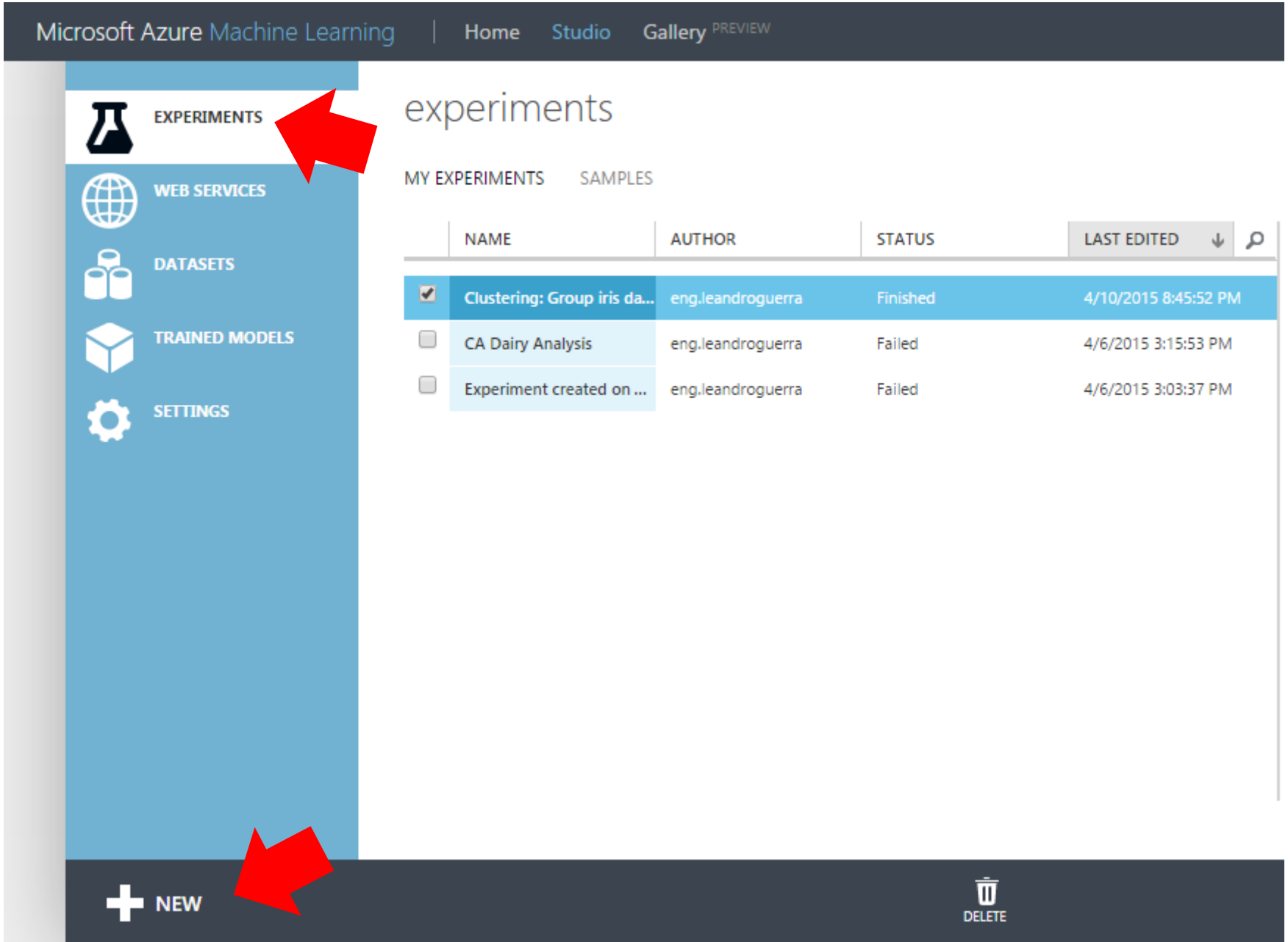
Windows Azure
Machine Learning



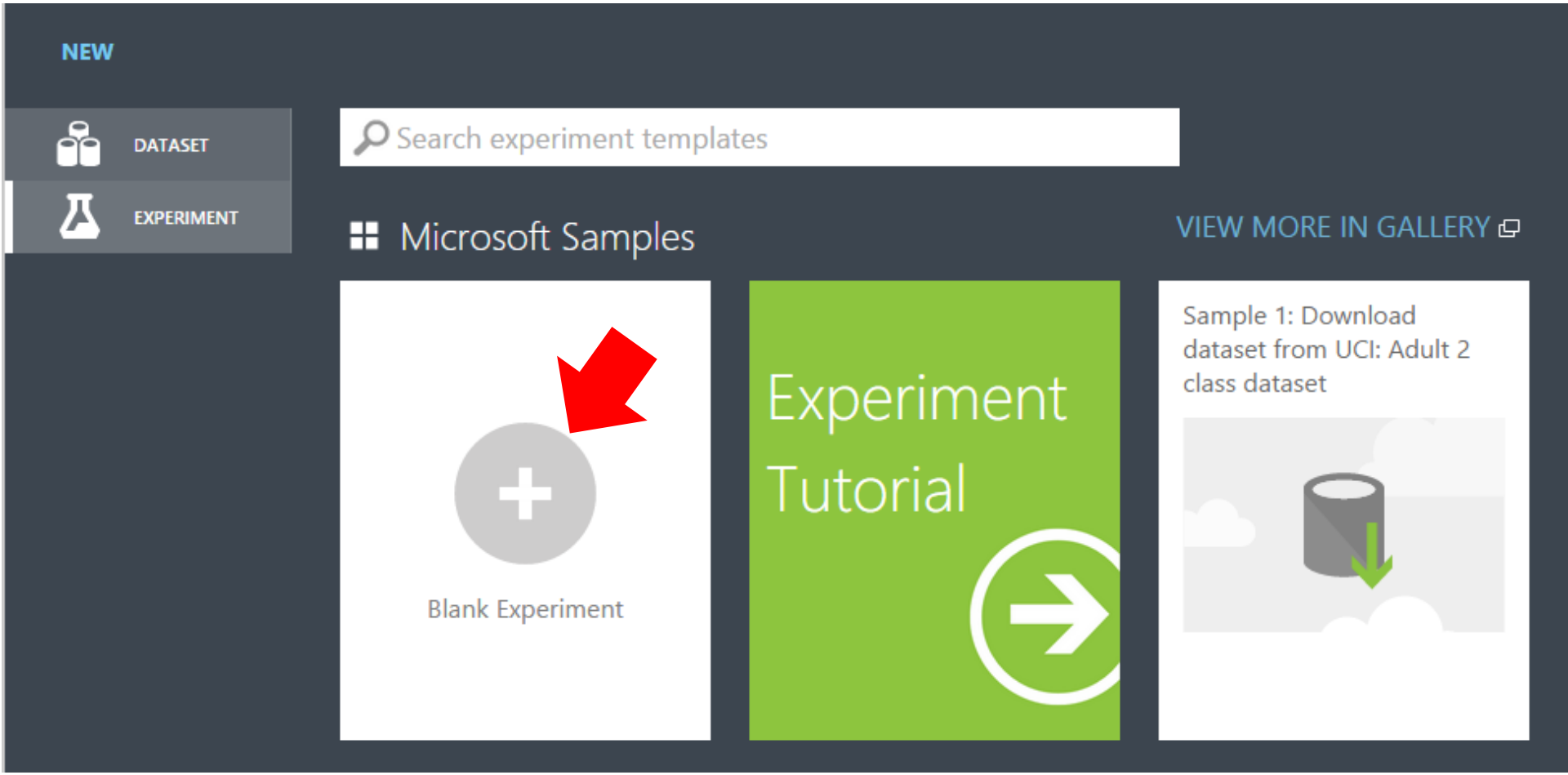
Windows Azure
Machine Learning



Windows Azure Machine Learning

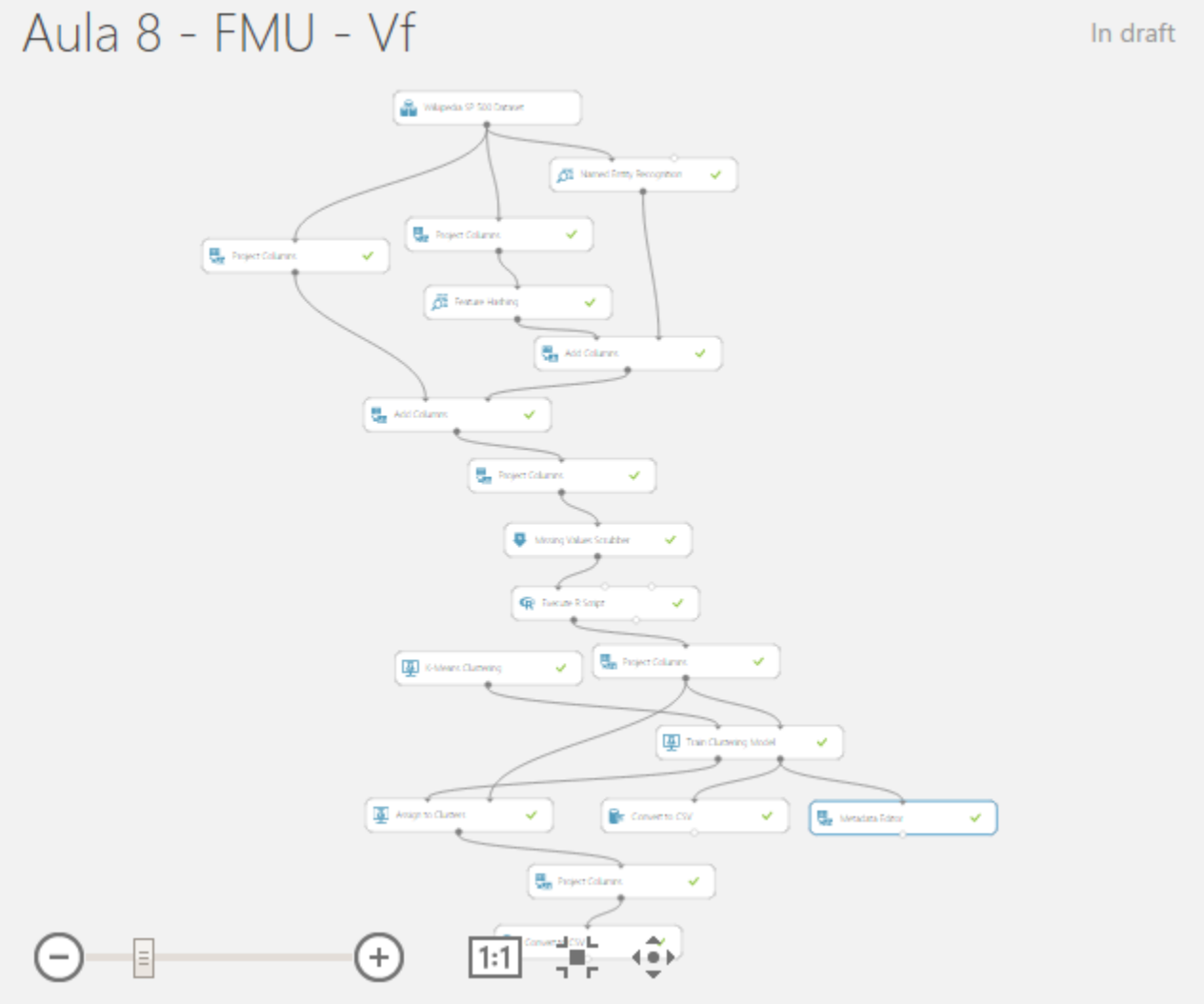


Windows Azure
Criando um novo Experimento

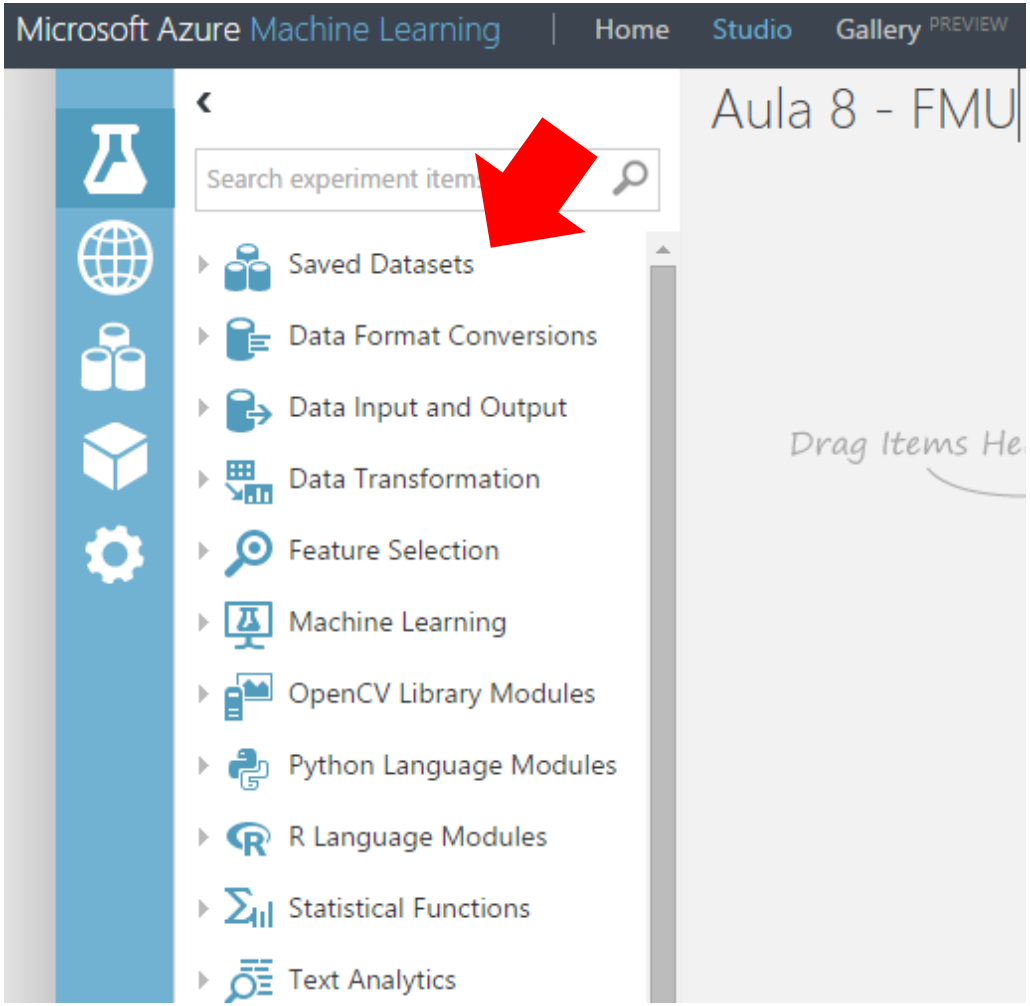


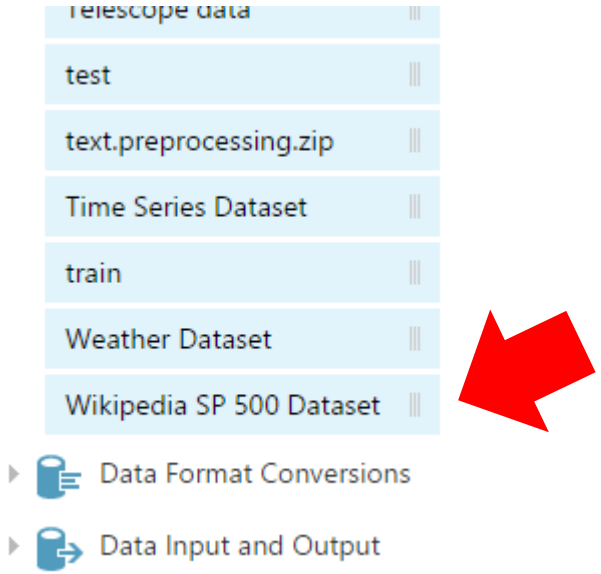
Windows Azure

Criando um novo Experimento – Nosso Objetivo

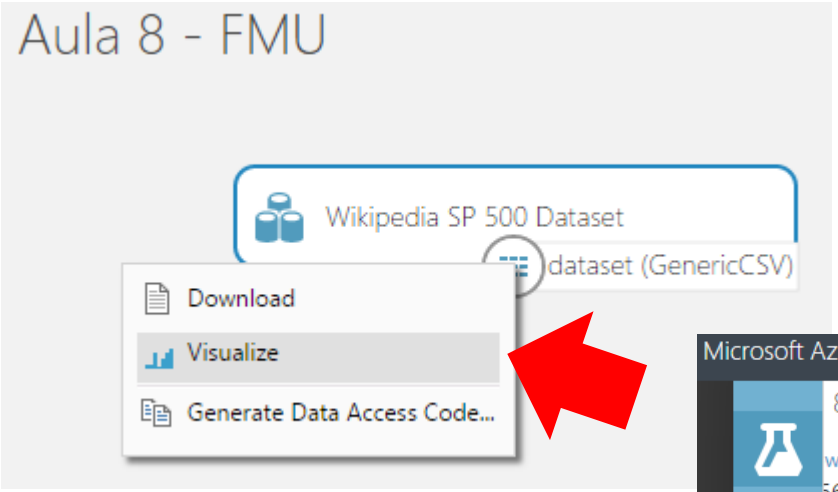


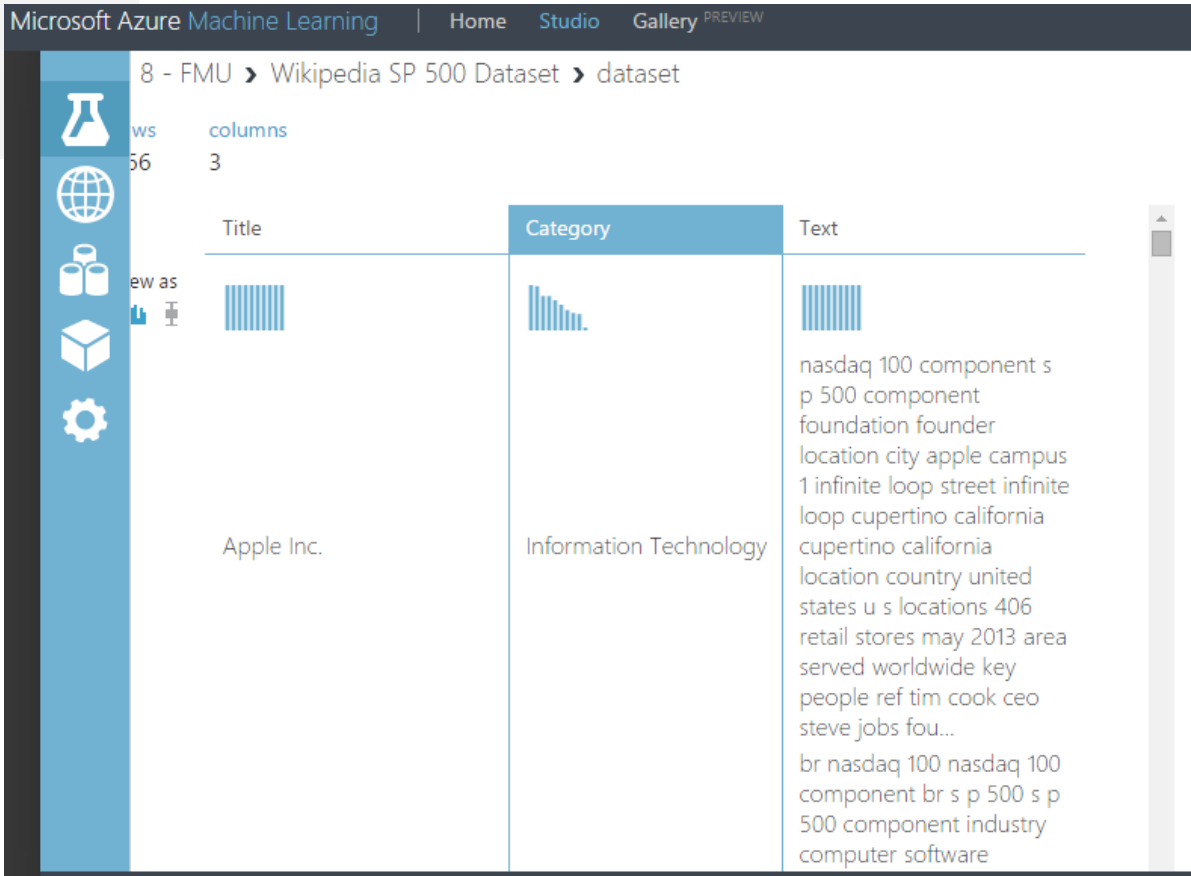
Windows Azure
Escolha do dataset e visualização



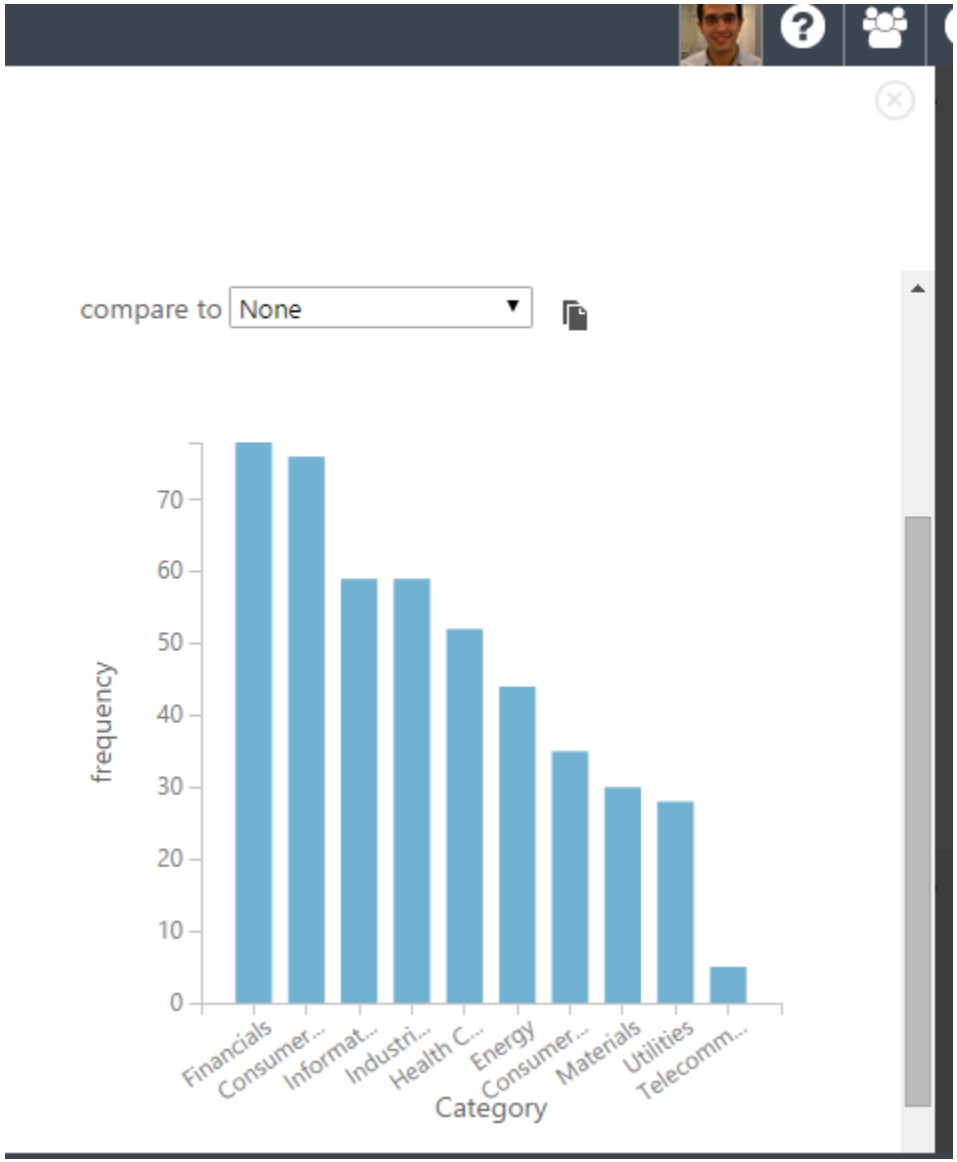


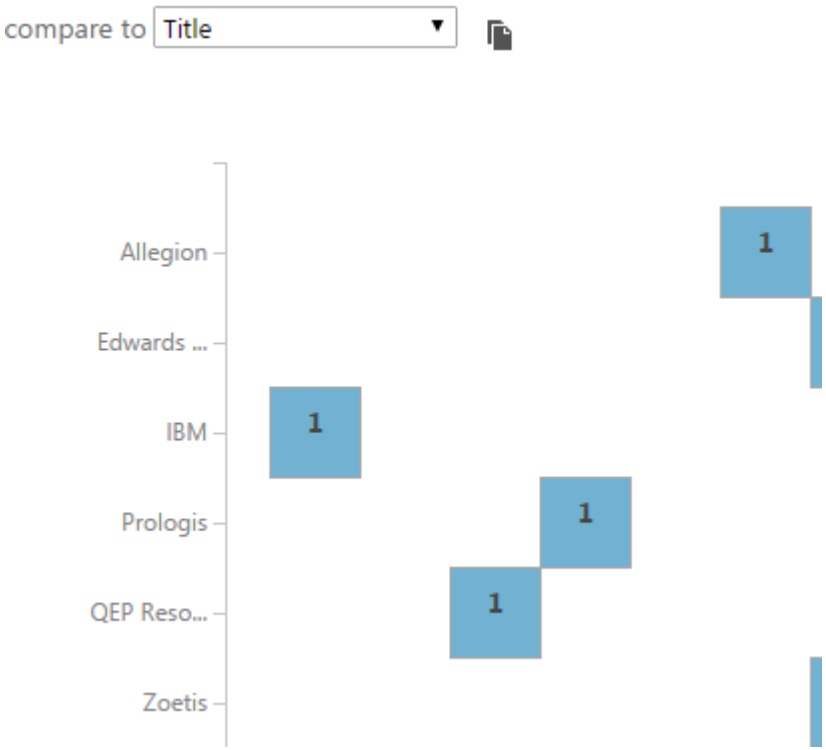
Windows Azure
Visualização dos Dados



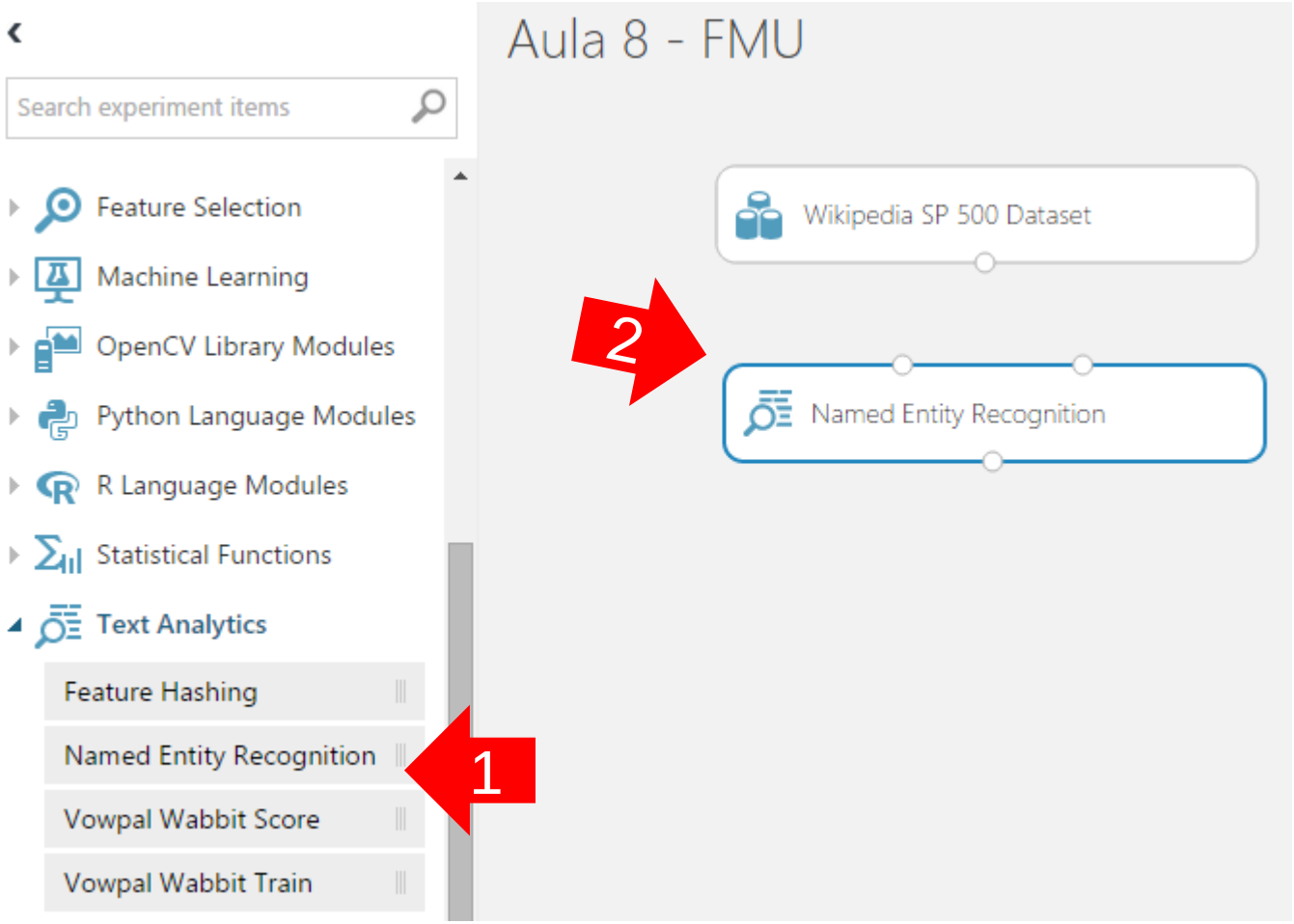


Windows Azure
Visualização dos Dados

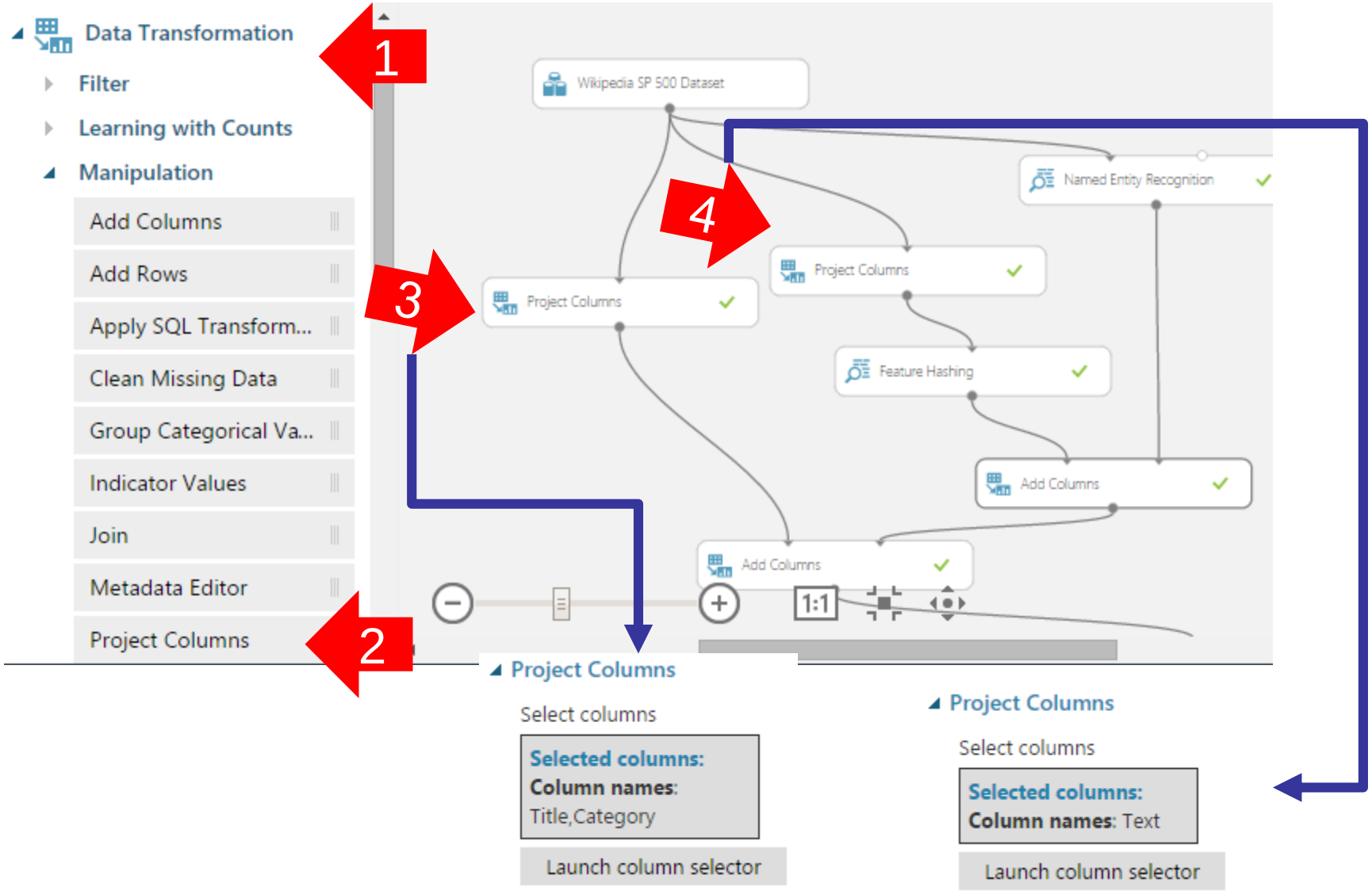




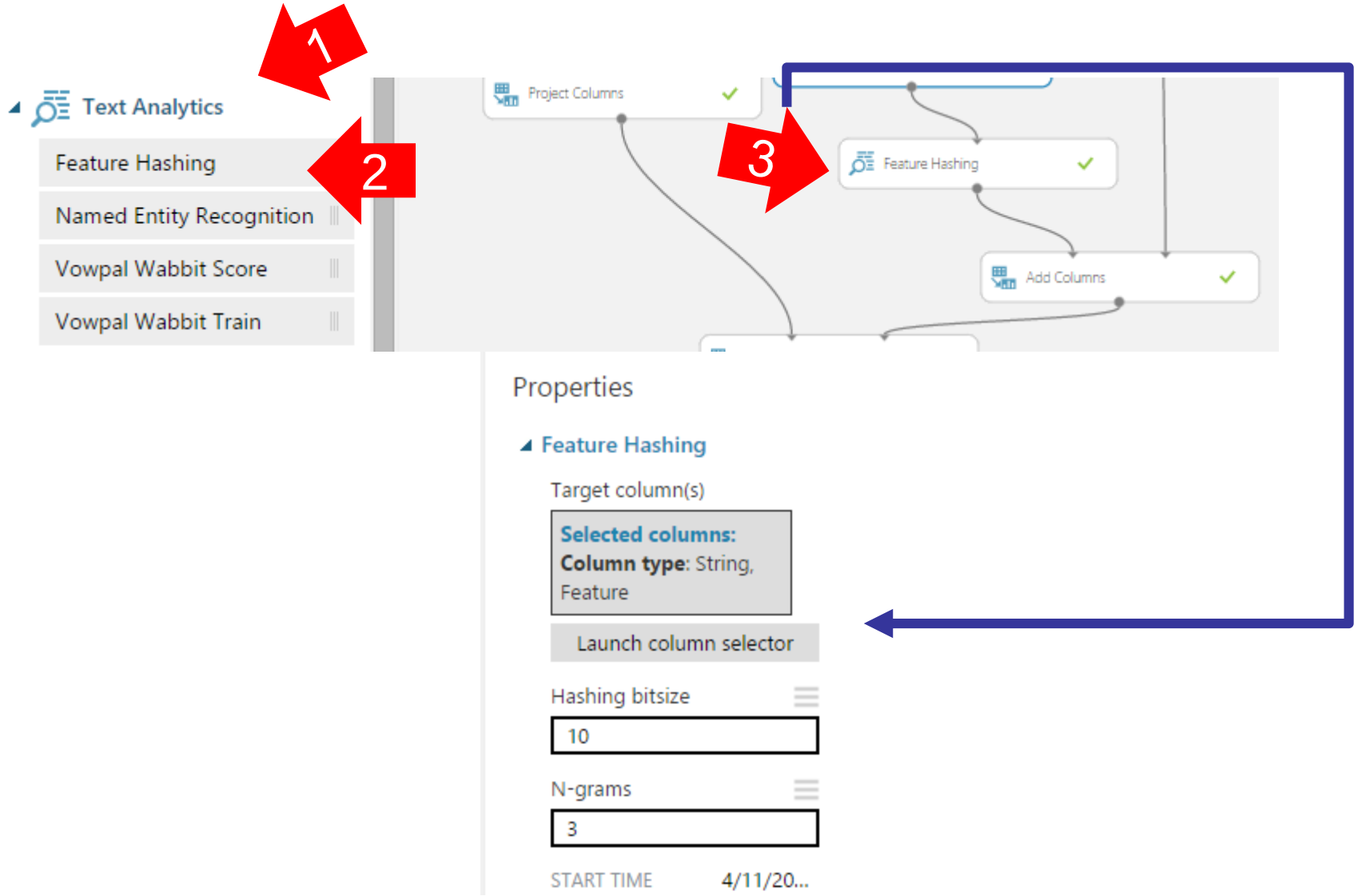
Windows Azure
Classificação de Empresas



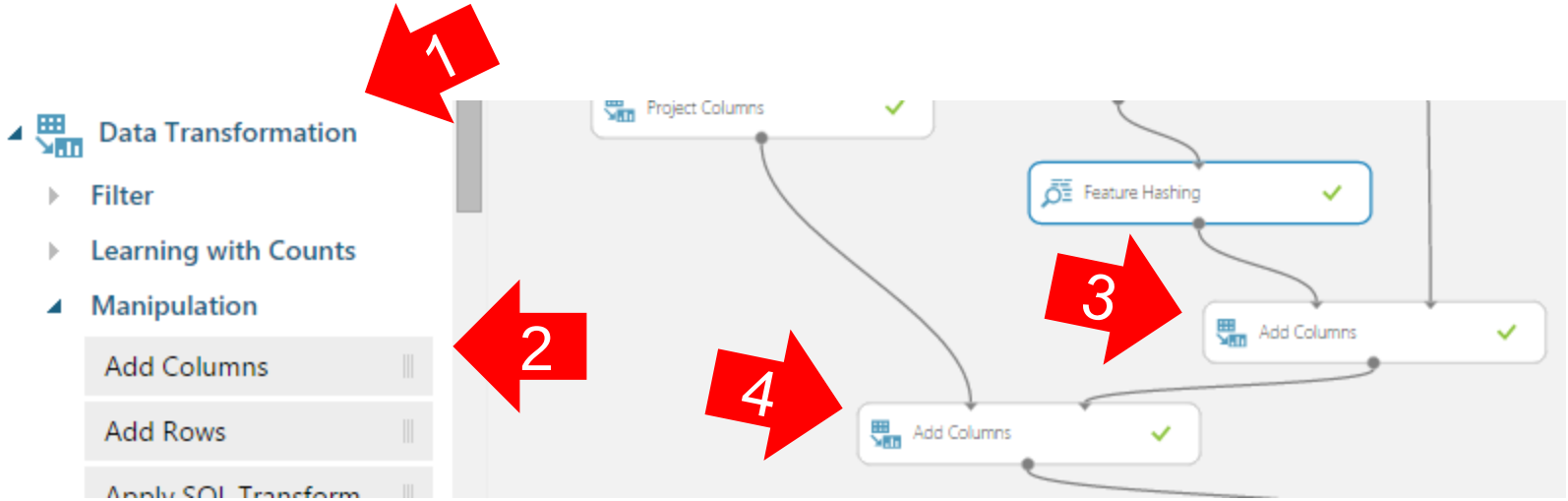
Windows Azure
Criando a Estrutura



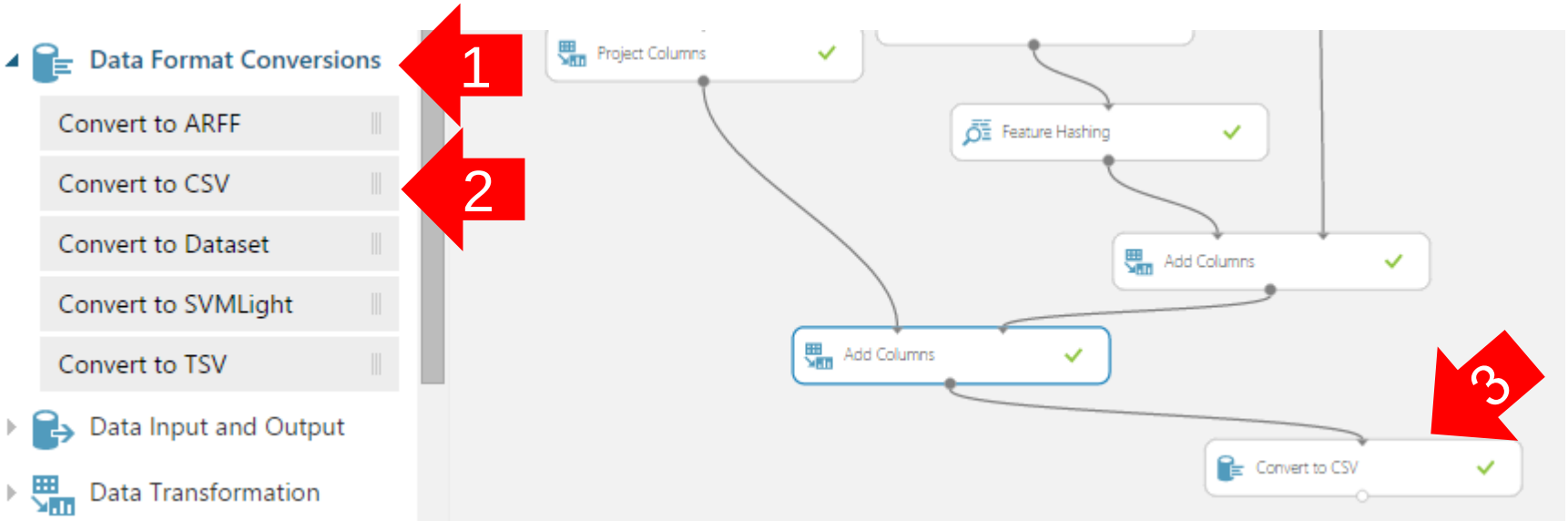
Windows Azure
Criando a Estrutura



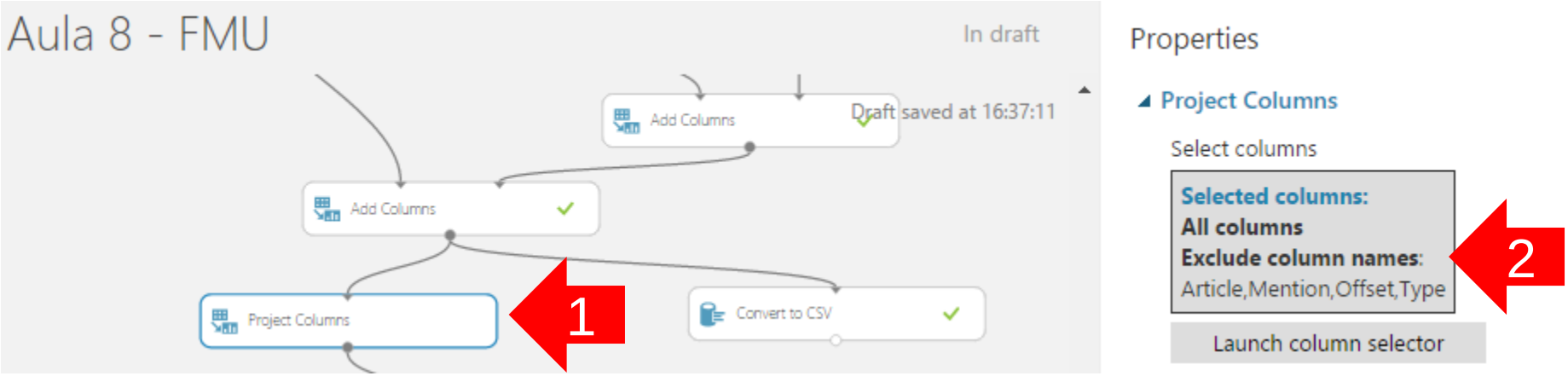
Windows Azure
Criando a Estrutura

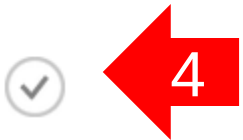


Windows Azure
Visualização parcial

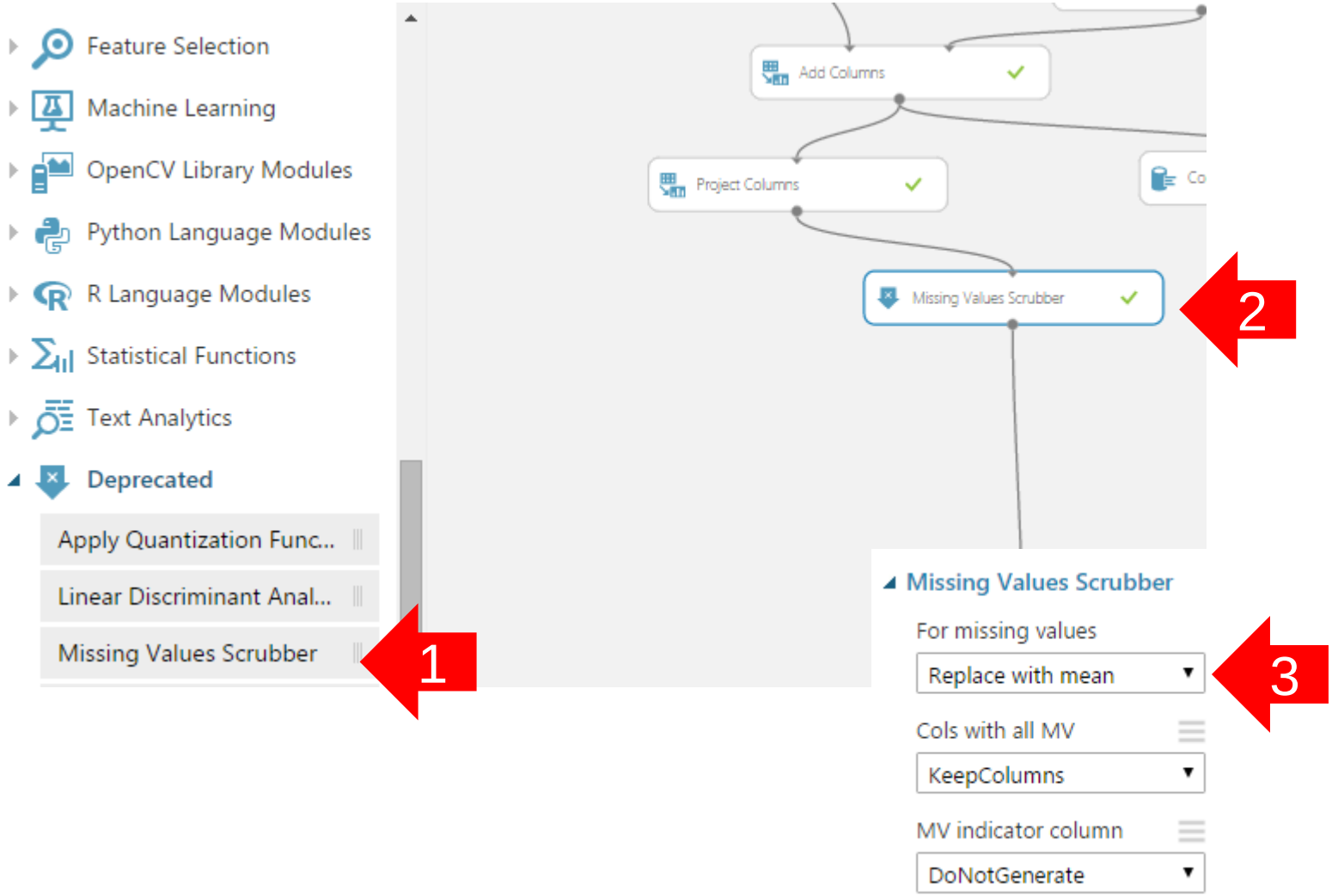


Windows Azure
Excluindo algumas colunas

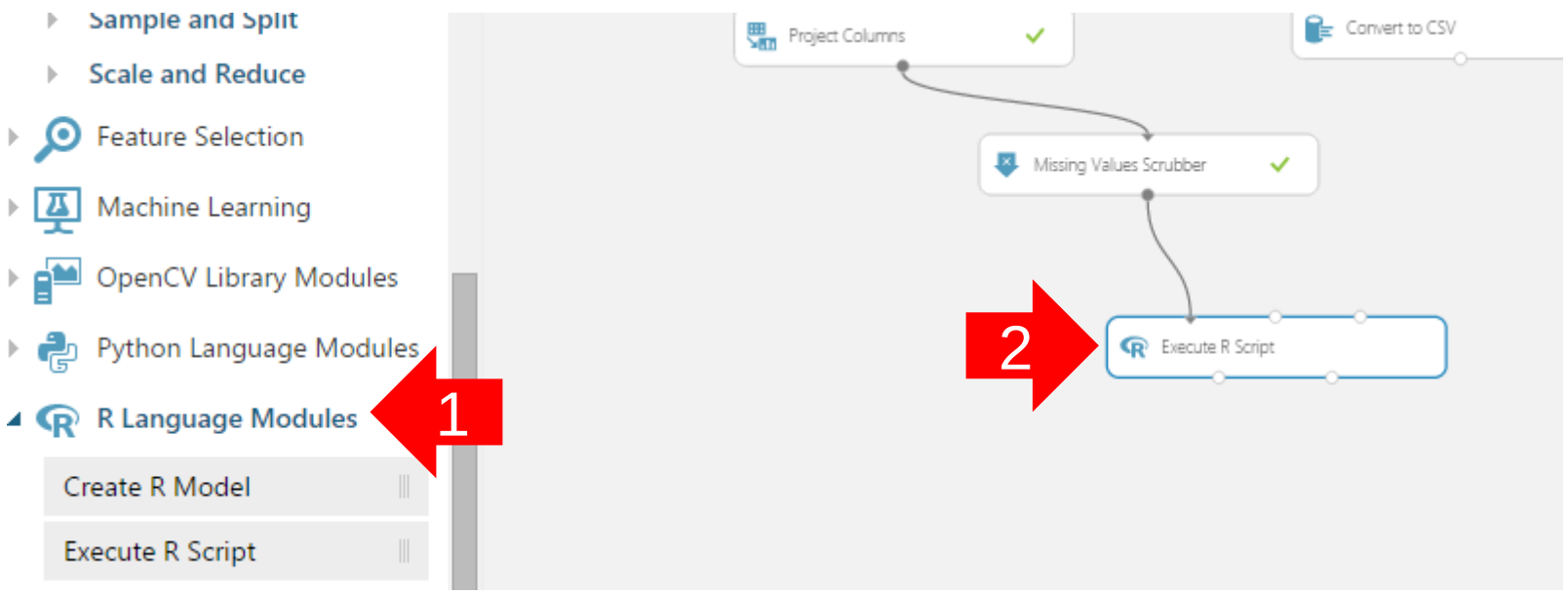




Windows Azure
Eliminando missing values



Windows Azure
Executando um script no R



Windows Azure
Executando um script no R

Execute R Script

R Script

```
# Inserindo os dados
base <- maml.mapInputPort(1) # class: data.frame

Title_Category = base[,1:2]

# Análise de Componentes Principais (PCA)
pca = prcomp(base[,4:1028])
top_pca = data.frame(pca$x[,1:10])
dataframe = cbind(Title_Category,top_pca)

# Visualizando a PCA
plot(pca)

# Enviado o dataframe para a saida
maml.mapOutputPort("dataframe");
```

Windows Azure

Executando um script no R - Resultado

8 - FMU › Execute R Script › Result Dataset

ws 56 | columns 12

ew as

Title	Category	PC1	PC2	PC3	PC4
Apple Inc.	Information Technology	-1196.324169	-136.624916	19.972306	-79.8467
Adobe Systems	Information Technology	-266.248139	-47.893461	35.306568	56.53345
General Motors	Consumer Discretionary	-568.967279	20.830032	-48.233082	21.780172
General Electric	Energy	-394.76521	11.138154	-7.509128	0.695133
Harley-Davidson	Consumer Discretionary	-796.350029	-24.003269	-87.089201	-189.3293
Intel	Information Technology	-1031.574792	-42.587229	29.673068	-35.15889
Microsoft	Information Technology	-510.561424	-42.020697	18.699546	-6.184506
Mattel	Consumer Discretionary	107.893397	1.399807	-8.120238	-10.58427

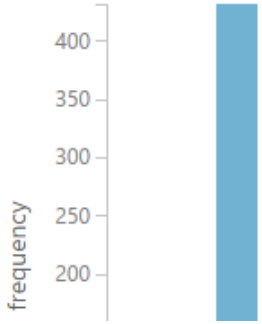
Statistics

Visualizations

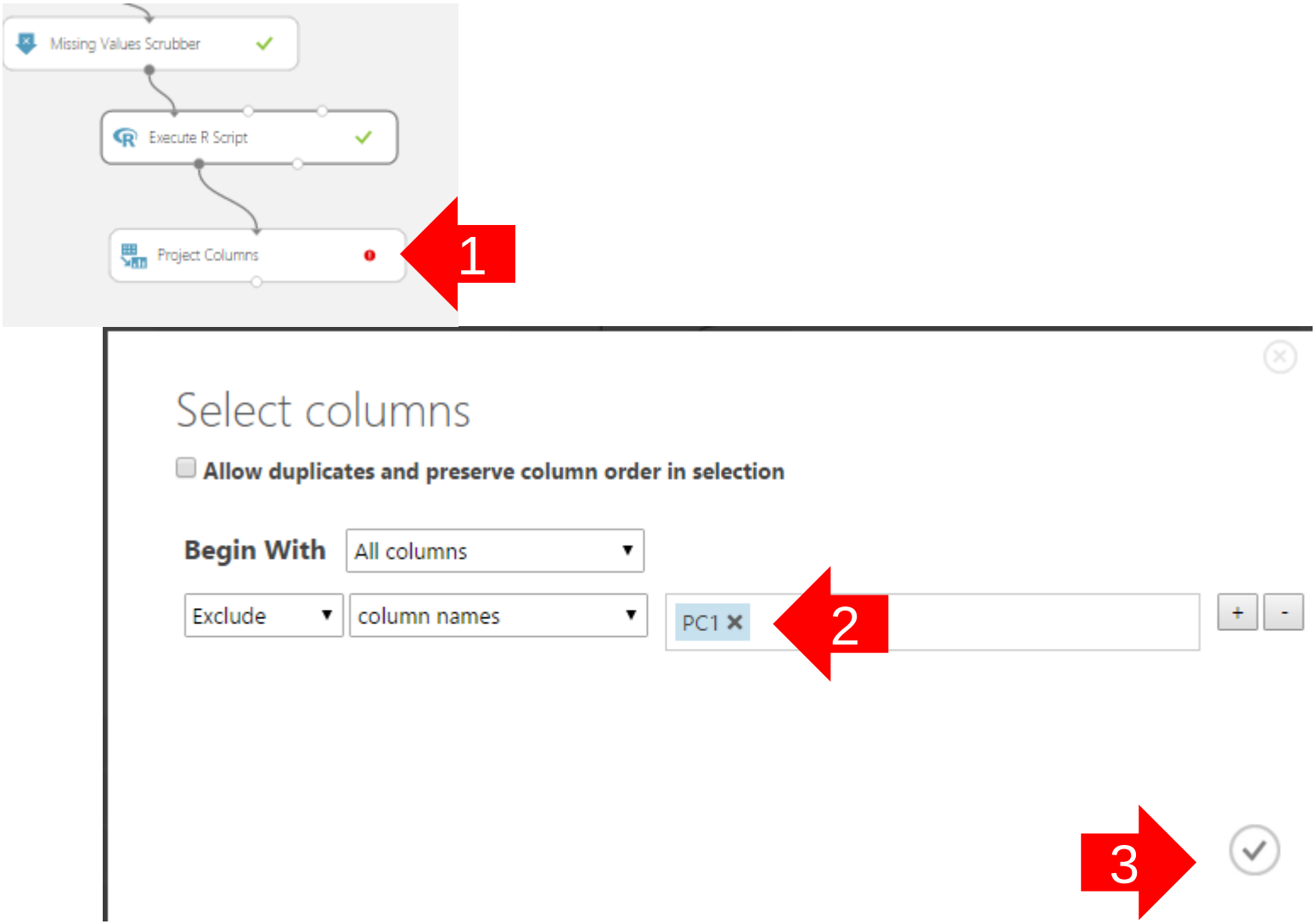
PC2

Histogram

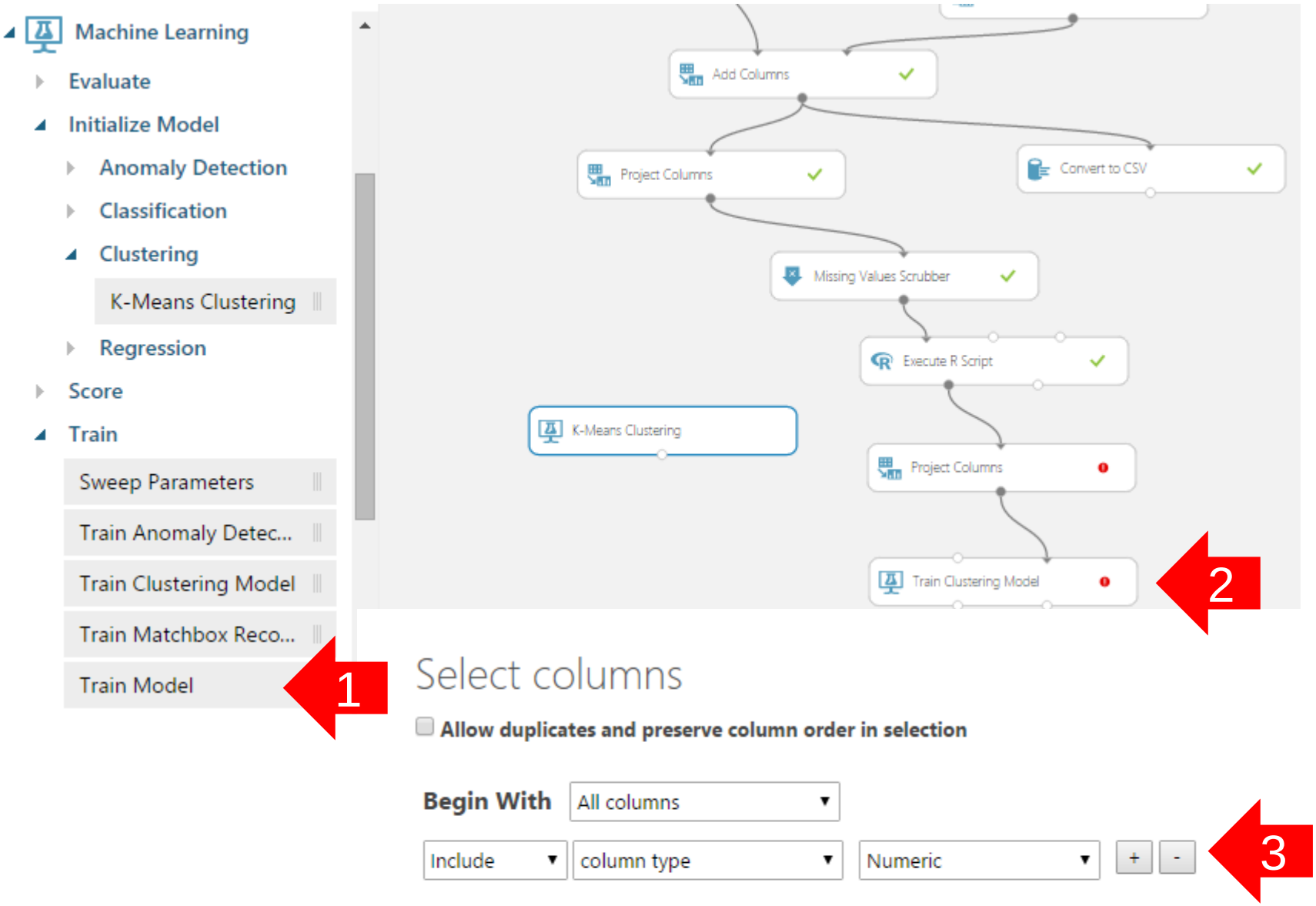
compare to None



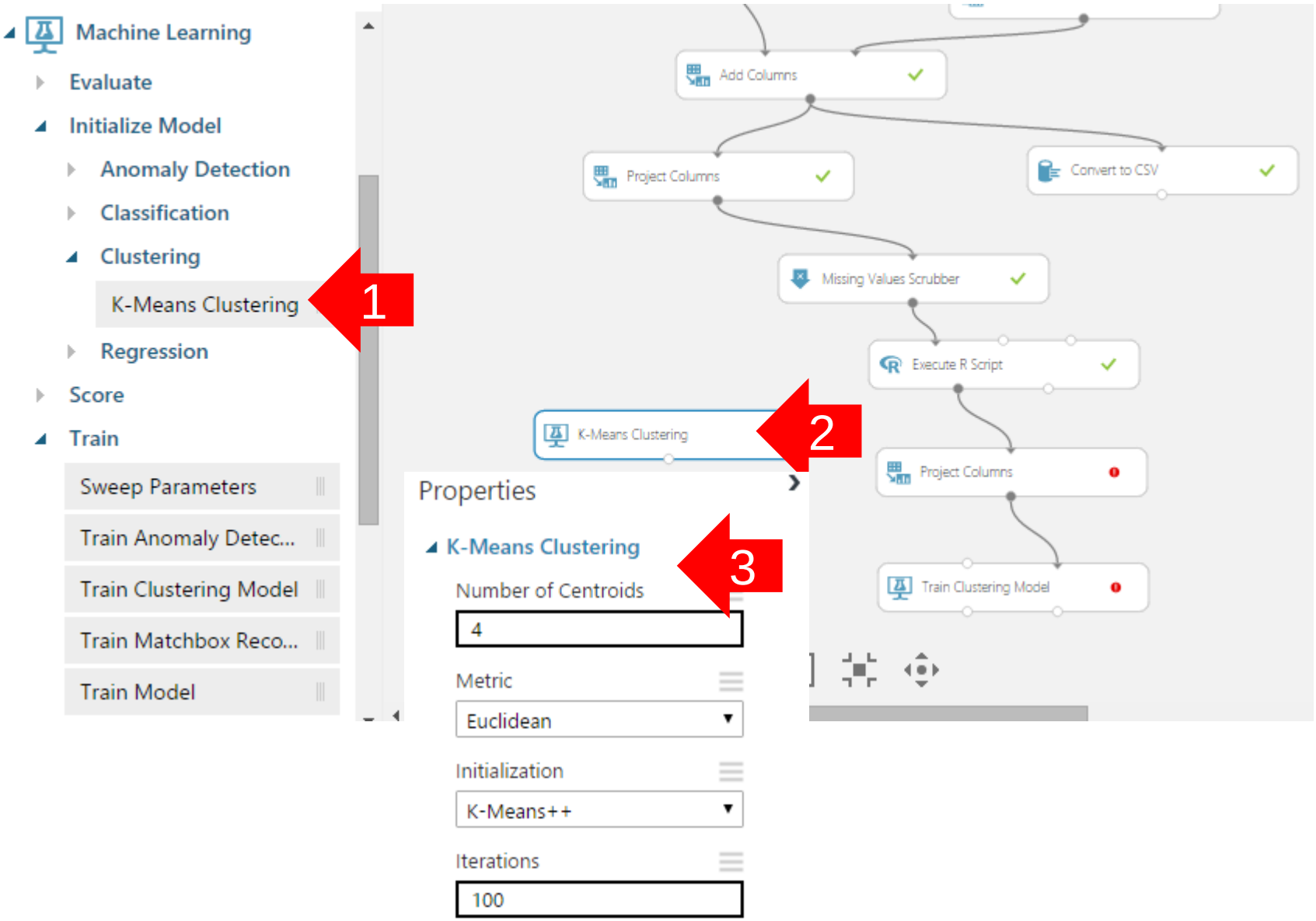
Windows Azure
K-Means



Windows Azure
K-Means

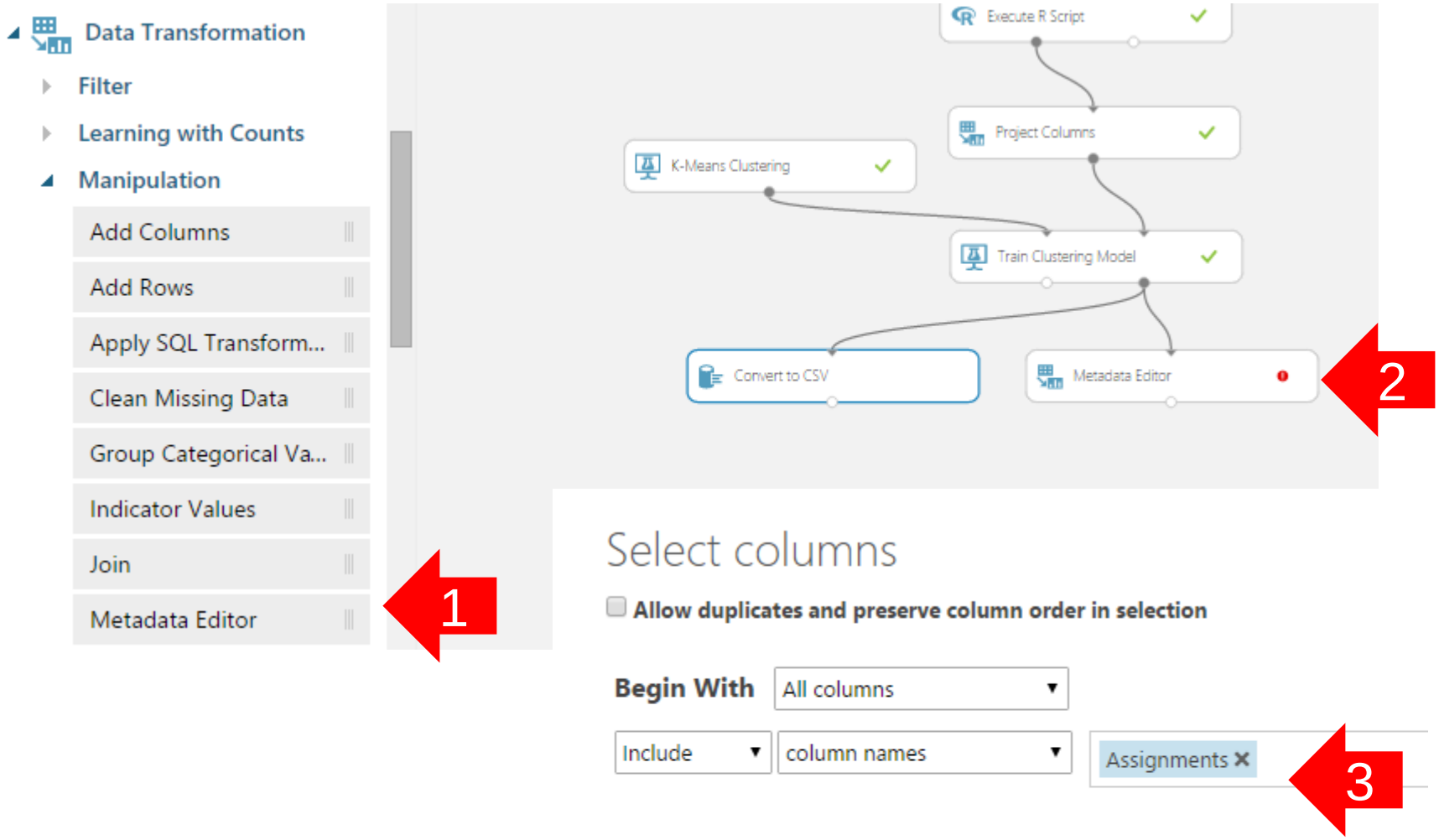


Windows Azure
K-Means

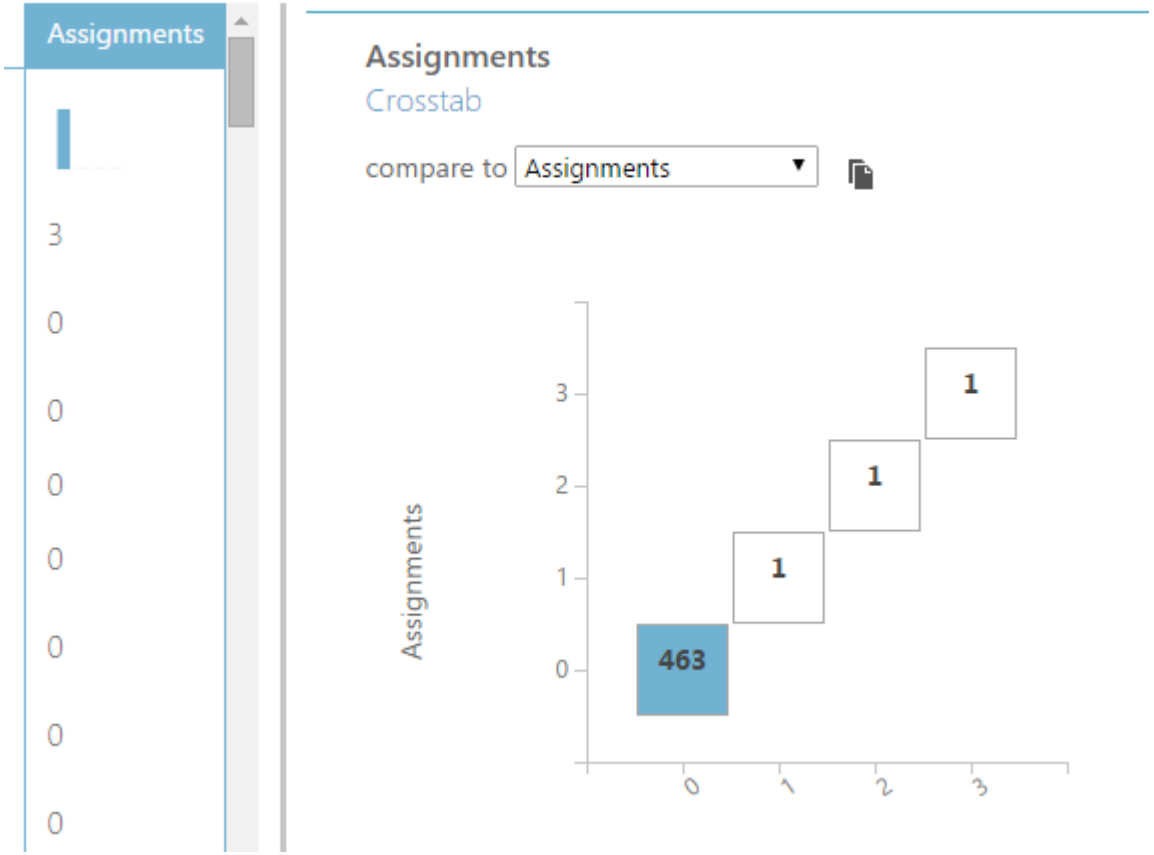


Windows Azure

K-Means - Resultados

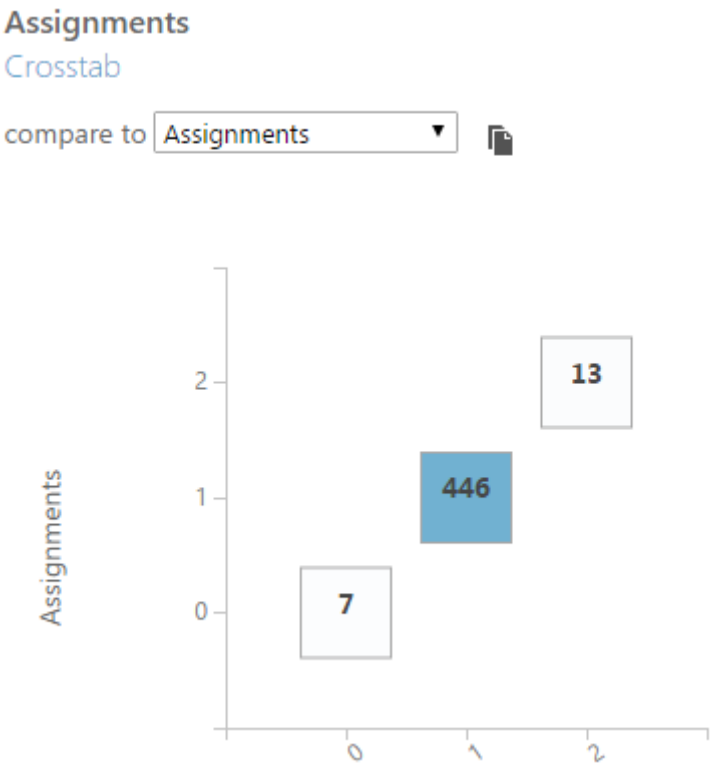


Windows Azure
K-Means - Resultados



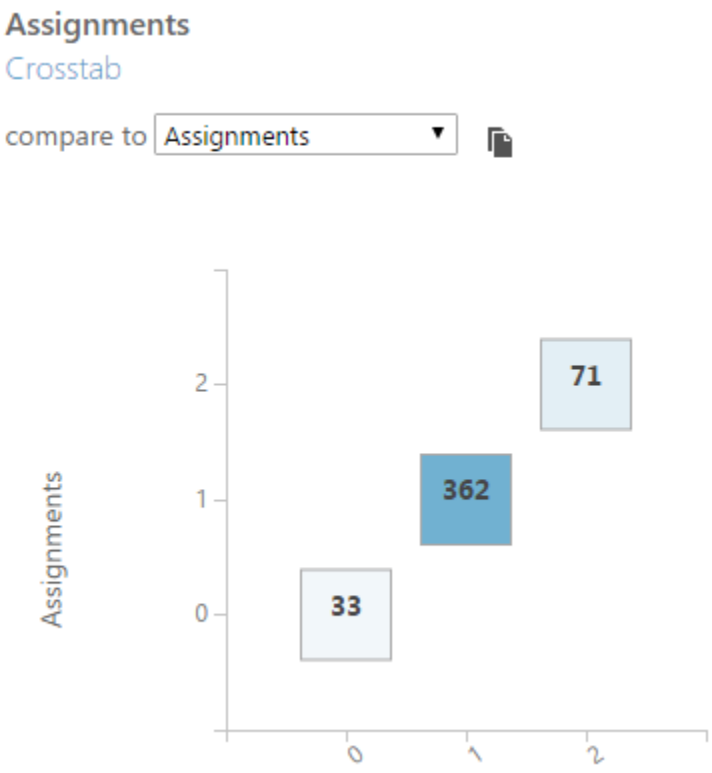
Windows Azure

K-Means – Resultados – Com Bigramas e 3 Clusters



Windows Azure

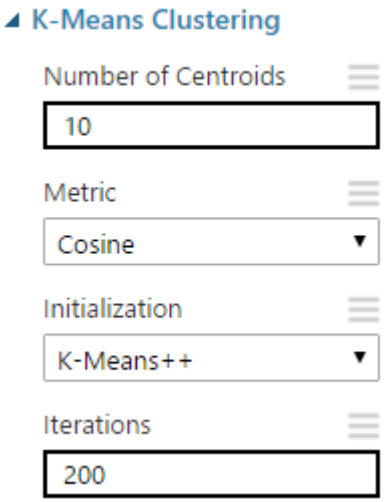
K-Means – Resultados – Com Hash de 15 bitsize

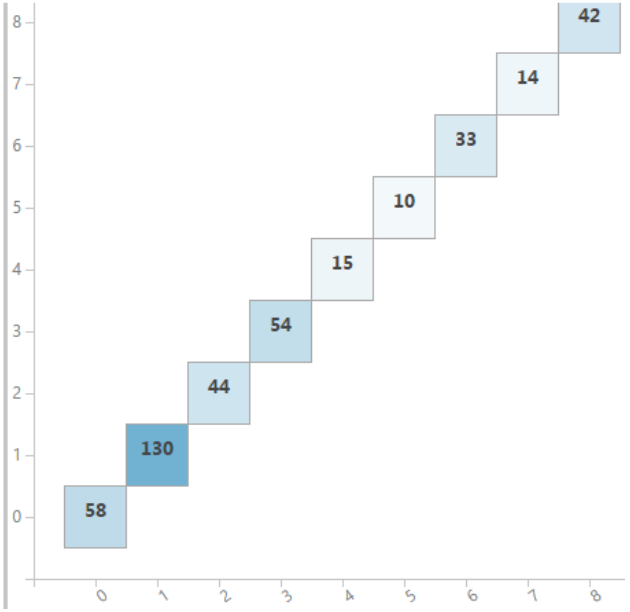


Windows Azure
Configuração de melhor resultado

Hash com bitsize de 15

Configuração do K-Means





Windows Azure

K-Means – Resultados – Com Hash de 24 bitsize

2^24 = 16.777.216 de variáveis!

Execute R Script Error

Record start time:UTC 04/12/2015 00:26:39

Record end time: UTC 04/12/2015 01:20:27

Error message:

```
Unhandled Exception: OutOfMemoryException.
```

Business Intelligence