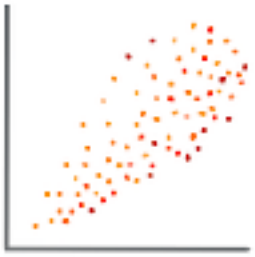
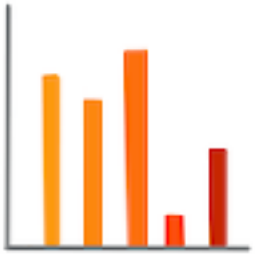
Aprendizaje no supervisado

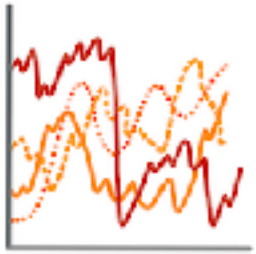
DATA VISUALIZATION



SCATTERPLOT

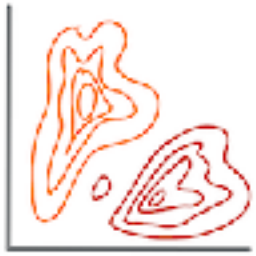


HISTOGRAM

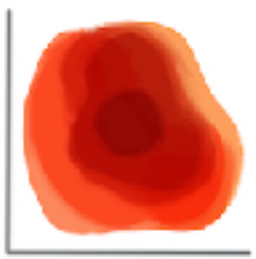


LINEPLOT

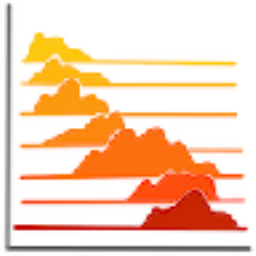
VIOLINPLOT



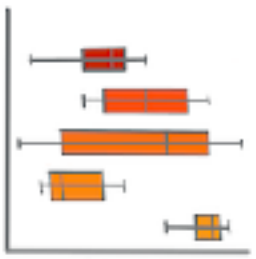
KDE PLOT



HEATMAP



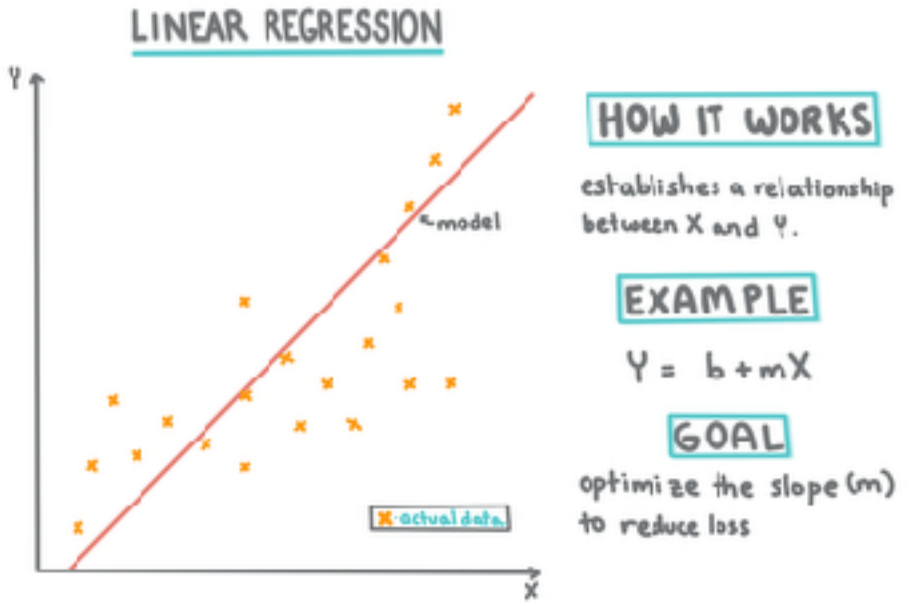
RIDGEPLOT

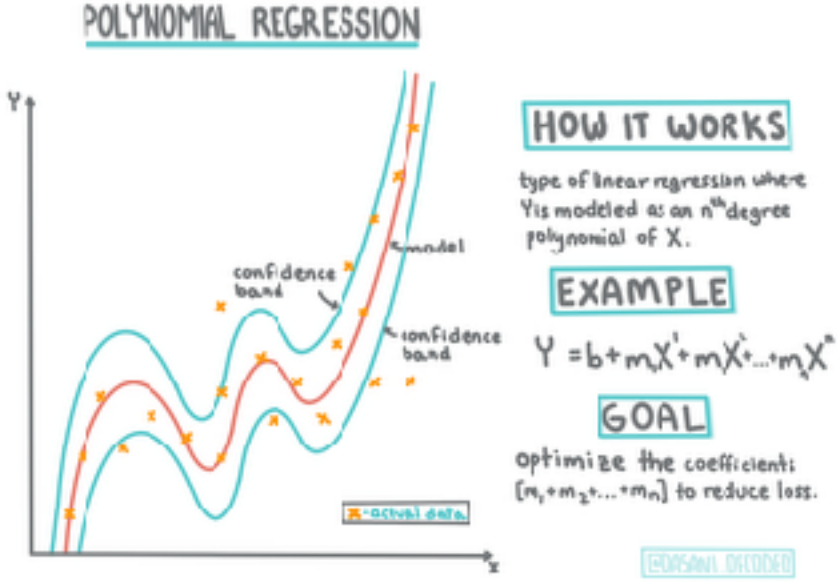


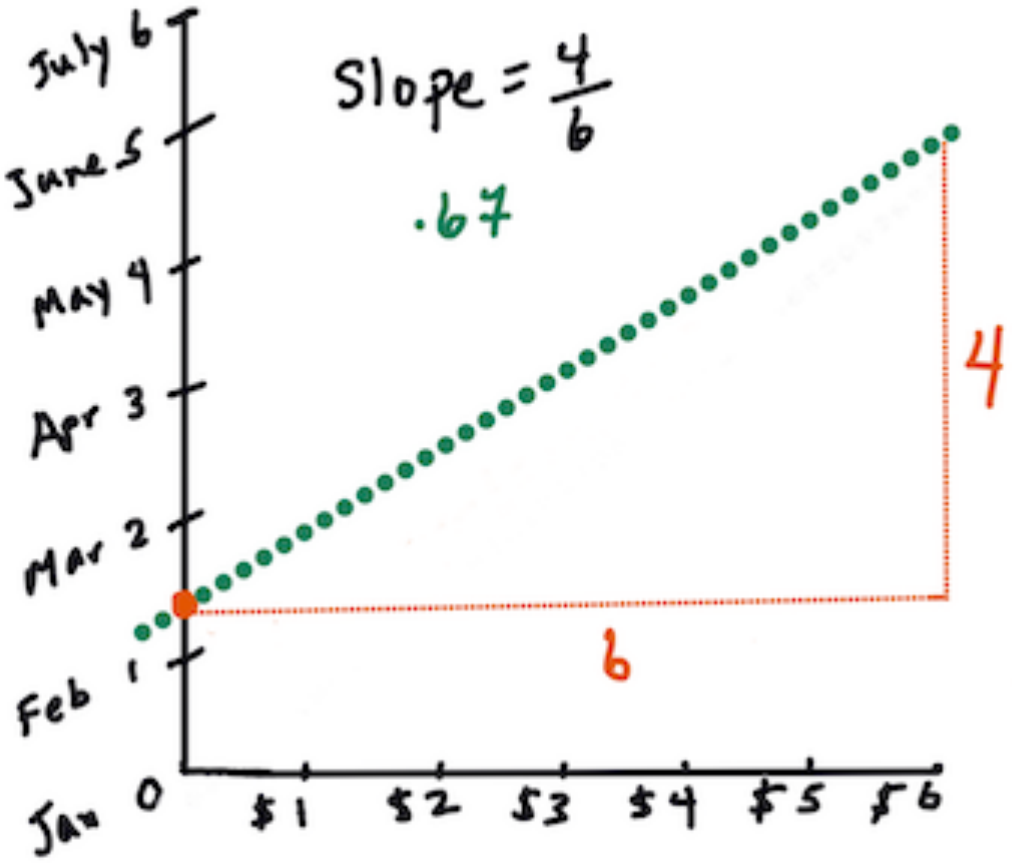
BOXPLOT

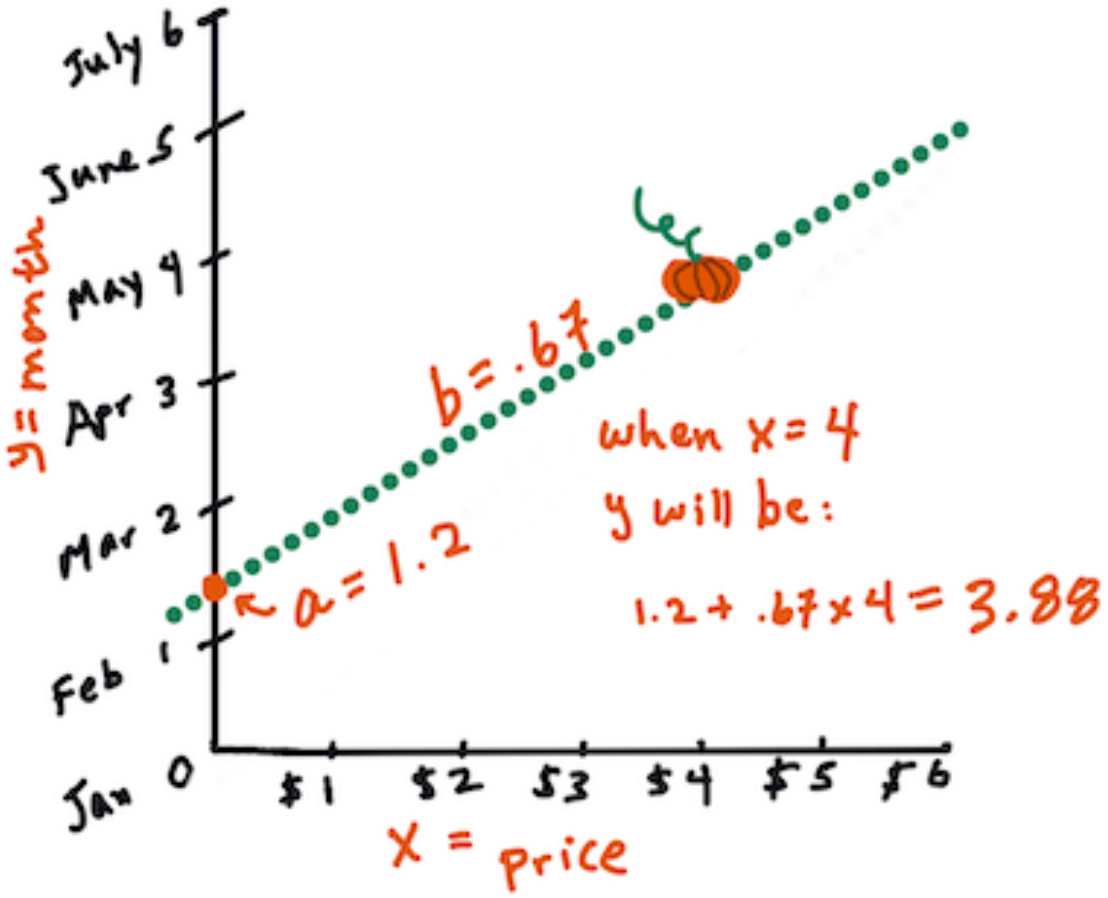
How you visualize data influences how you frame the problem

Regresiones





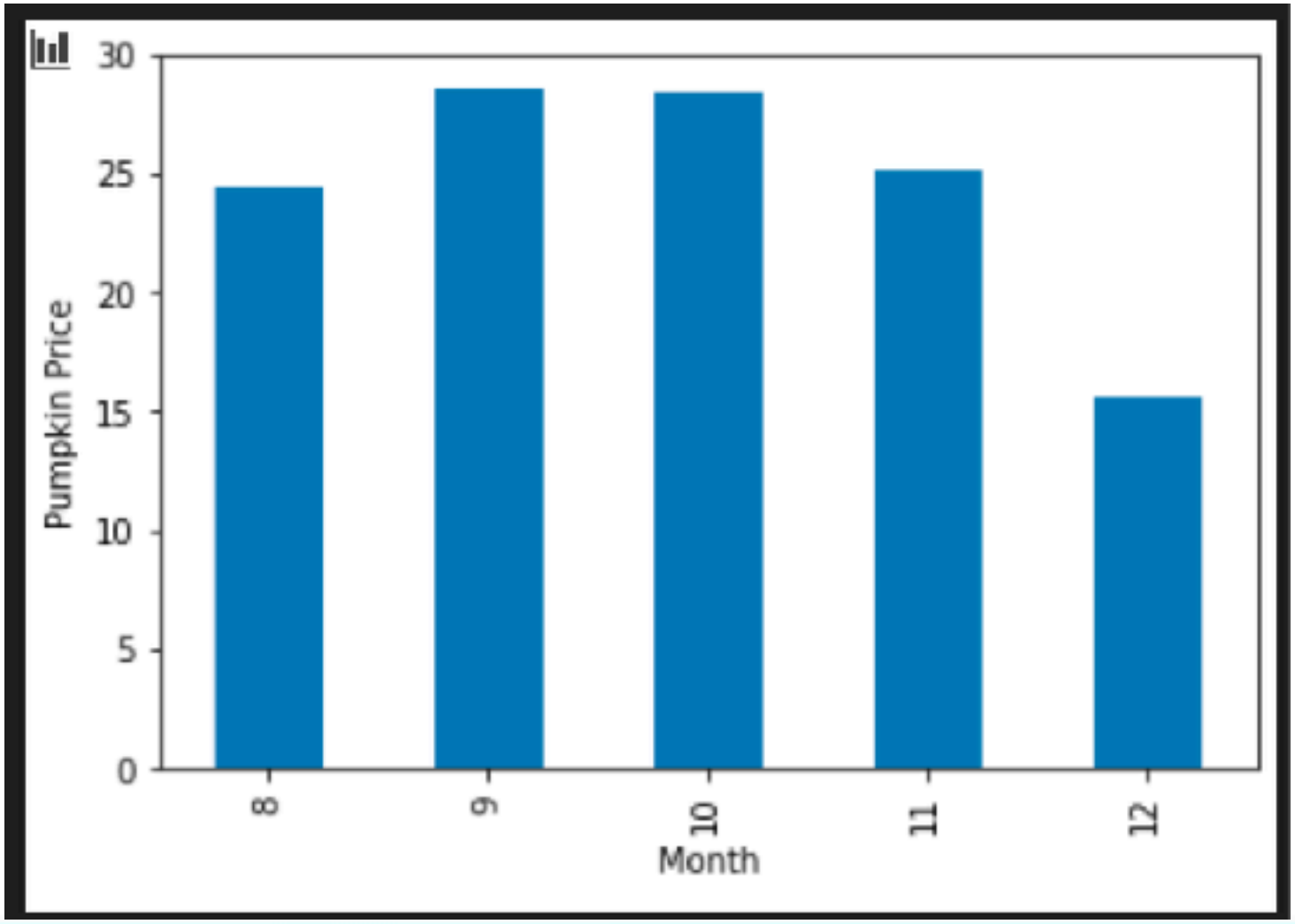

July 6
June 5
May 4
Apr 3
Mar 2
Feb 1
Jan 0
$1
$2
$3
$4
$5
$6
Slope = 4/6
.67
4
6


July 6
June 5
May 4
Apr 3
Mar 2
Feb 1
Jan 0
y = month
$1
$2
$3
$4
$5
$6
X = price
b = .67
a = 1.2
when x = 4
y will be:
1.2 + .67 x 4 = 3.88

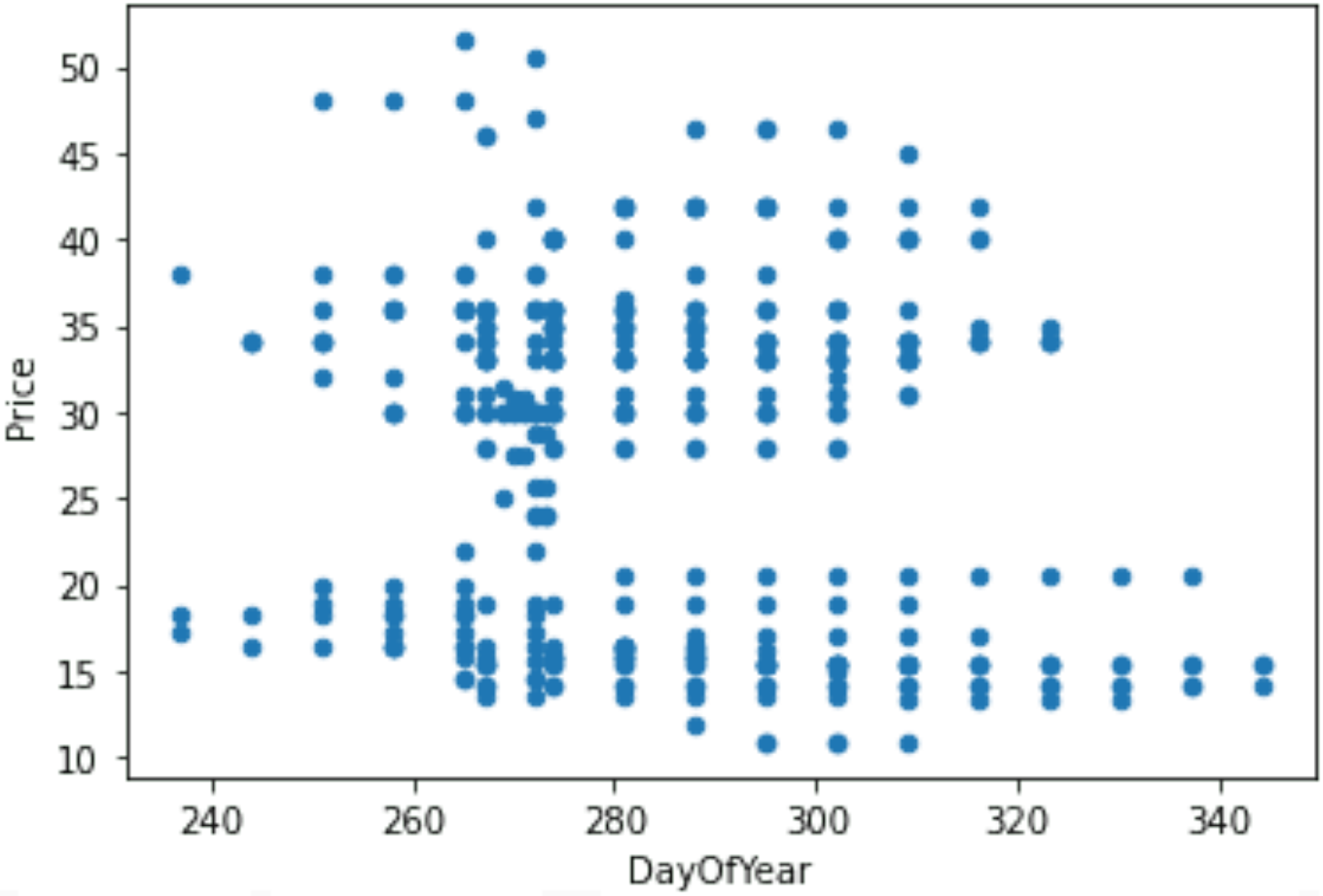
Ejemplo

ID	Month	DayOfYear	Variety	City	Package	Low Price	High Price	Price
70	9	267	PIE TYPE	BALTIMORE	1 1/9 bushel cartons	15.0	15.0	13.636364
71	9	267	PIE TYPE	BALTIMORE	1 1/9 bushel cartons	18.0	18.0	16.363636
72	10	274	PIE TYPE	BALTIMORE	1 1/9 bushel cartons	18.0	18.0	16.363636
73	10	274	PIE TYPE	BALTIMORE	1 1/9 bushel cartons	17.0	17.0	15.454545
74	10	281	PIE TYPE	BALTIMORE	1 1/9 bushel cartons	15.0	15.0	13.636364

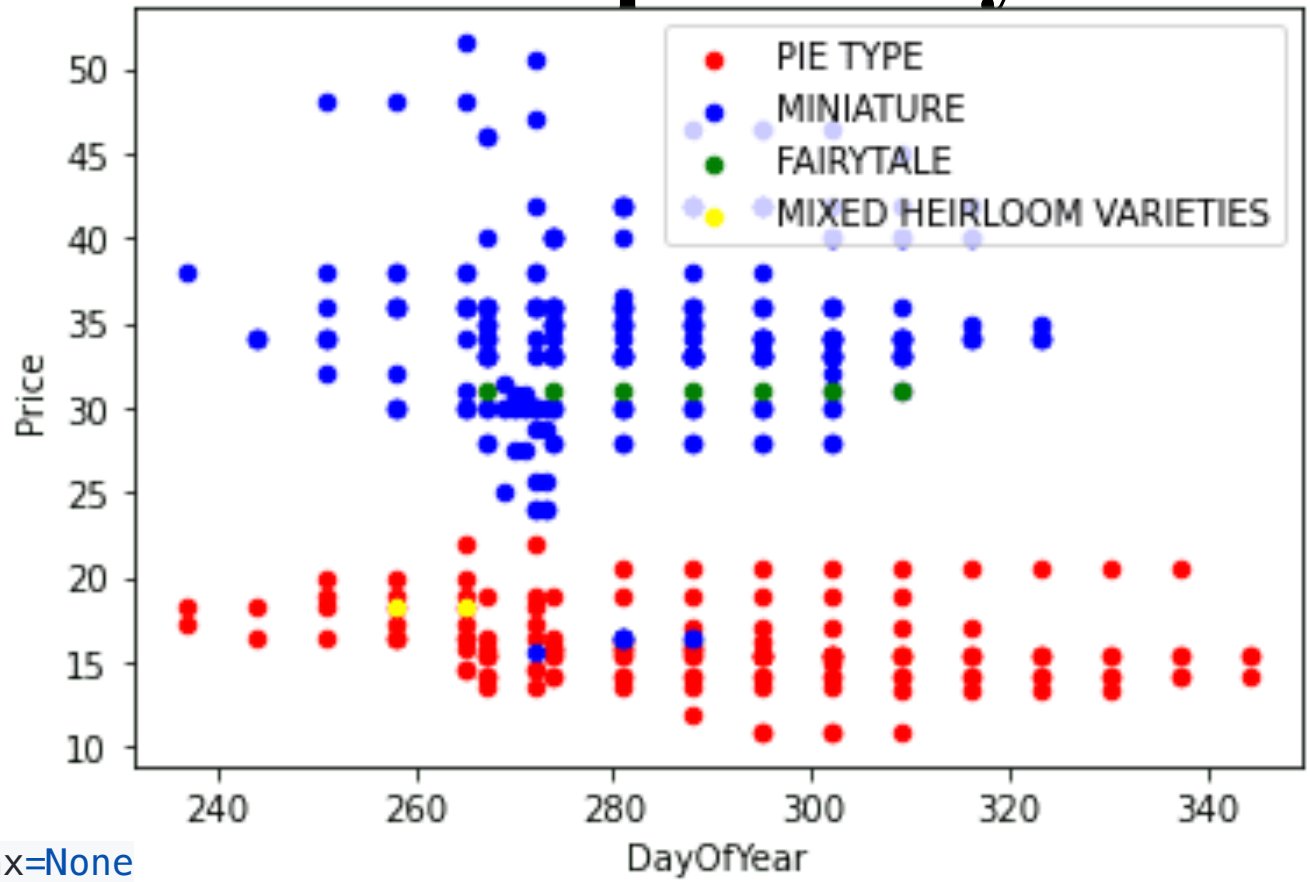
Precios promedio



Relación día del año - Precio

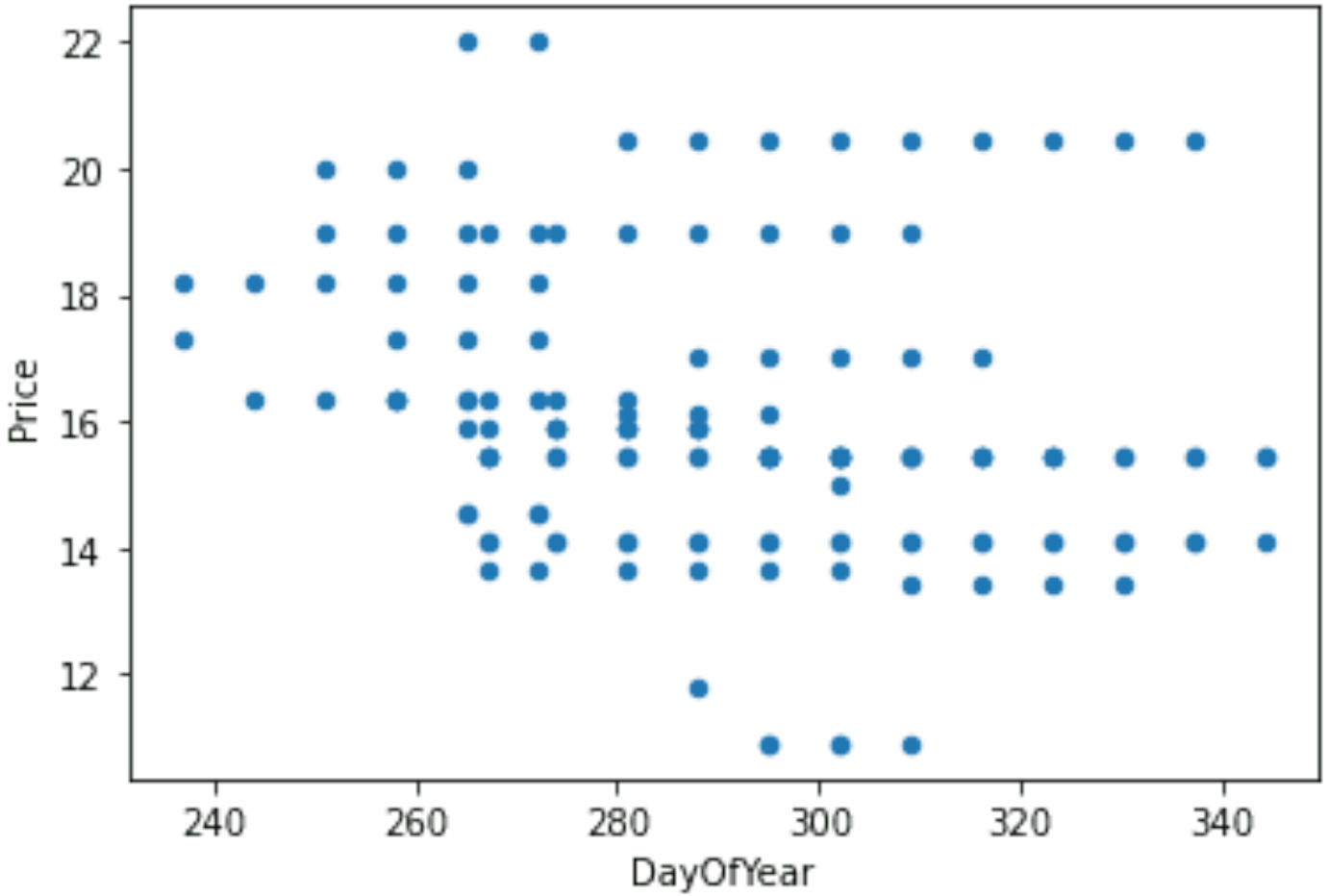


Relación entre precios y variedades



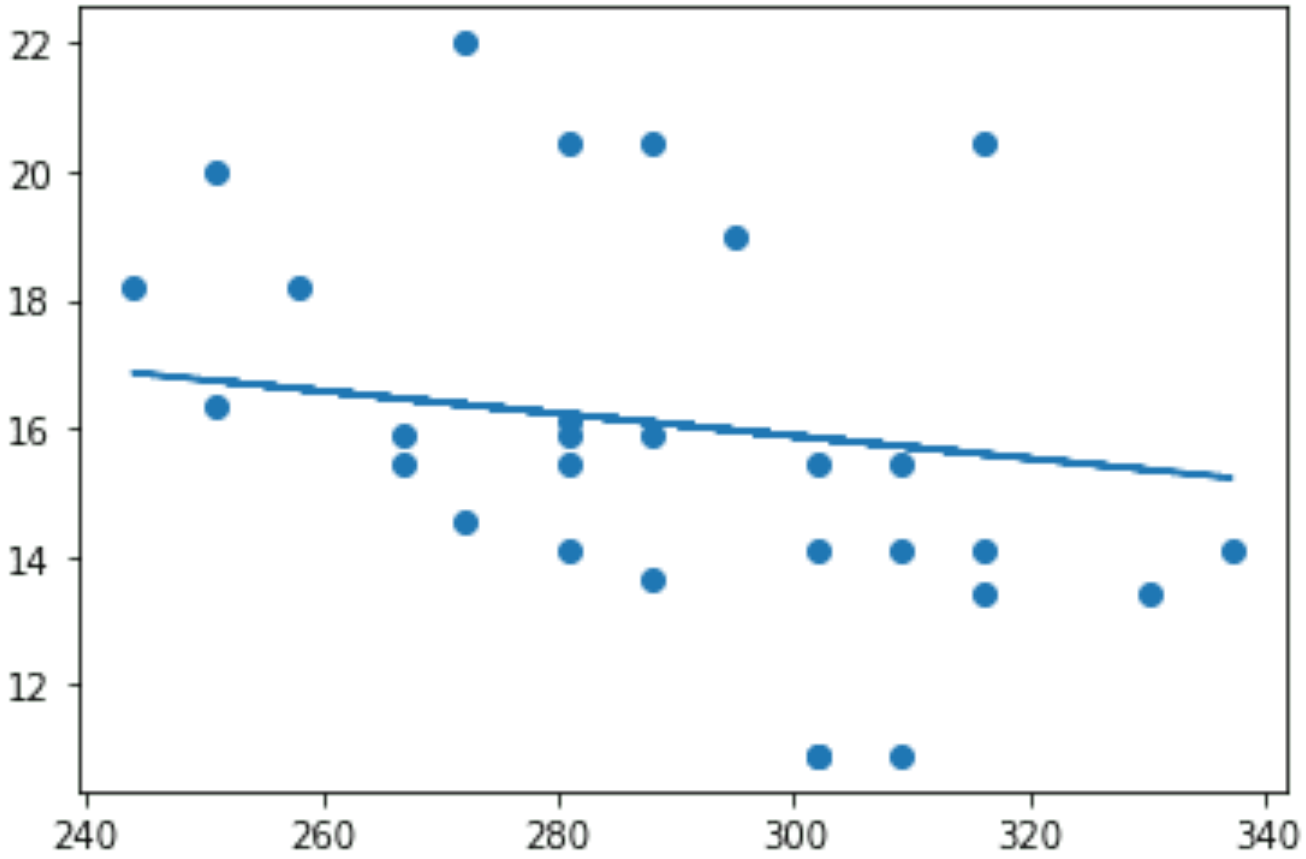

```
ax=None
colors = ['red','blue','green','yellow']
for i,var in enumerate(new_pumpkins['Variety'].unique()):
    df = new_pumpkins[new_pumpkins['Variety']==var]
    ax = df.plot.scatter('DayOfYear','Price',ax=ax,c=colors[i],label=var)
```

Foco en una variedad



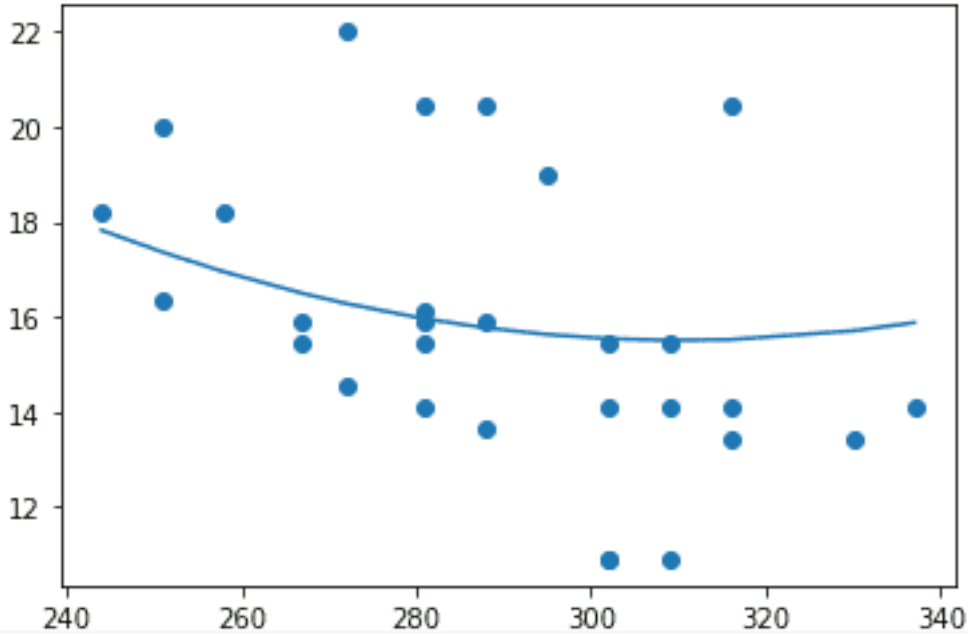

```
pie_pumpkins = new_pumpkins[new_pumpkins['Variety']=='PIE TYPE']
pie_pumpkins.plot.scatter('DayOfYear','Price')
```

Regresión lineal




```
lin_reg = LinearRegression()
lin_reg.fit(X_train,y_train)
```

Regresión polinomial




```
from sklearn.preprocessing import PolynomialFeatures
from sklearn.pipeline import make_pipeline

pipeline = make_pipeline(PolynomialFeatures(2), LinearRegression())

pipeline.fit(X_train,y_train)
```

Categorías por características