5G00ET68-3001 Data-analyysi ja tekoälyn perusteet

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

x = pd.Series([1,2,3,4,5,6,7,8])
y = 2 * x + 3

df = pd.DataFrame({'x':x, 'y':y})

df.plot(kind='scatter', x='x', y='y')
plt.show()

#X = df.loc[:, ['x','y']]
X = df.iloc[:, [0]]
Y = df.iloc[:, [1]]

regr = LinearRegression()
regr.fit(X, Y)

coef = regr.coef_
inter = regr.intercept_

print(f'Suoran yhtälö: y = {coef[0][0]}x + {inter[0]}')

y_pred = regr.predict([[5]])

df.plot(kind='scatter', x='x', y='y')
plt.scatter(5, y_pred, color='red')
plt.plot(df.x, df.y)
plt.show()
```

5G00ET68-3001 Data-analyysi ja tekoälyn perusteet

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error
from sklearn.model_selection import train_test_split

df = pd.read_csv('salary.csv')

df.plot(kind='scatter', x='YearsExperience', y='Salary')
plt.show()

X = df.iloc[:, [0]]
y = df.iloc[:, [1]]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=0)

regr = LinearRegression()
regr.fit(X_train, y_train)

coef = regr.coef_
inter = regr.intercept_

print(f'Suoran yhtälö on: y = {coef}x + {inter}')

y_pred = regr.predict(X_test)

r2 = r2_score(y_test, y_pred)

mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)

print(f'r2: {r2}')
print(f'mae: {mae}')
print(f'mse: {mse}')
print(f'rmse: {rmse}')

plt.scatter(X_train, y_train, color='red')
plt.plot(X_test, y_pred, color='blue')
plt.show()

print(f'Uuden työntekijän palkka 7v kokemuksella on: {regr.predict([[7]])}')
```