PROGRAM FOR THE PROBLEM STATEMENTS

```
import pandas as pd
# Load the dataset
file_path = '/content/sample_data/Super_Store_data.csv'
df = pd.read_csv(file_path,encoding='latin1')

# Display the first few rows of the dataset
print(df.head())
# Check for missing values
print(df.isnull().sum())

# Handle missing values (if any)
df = df.dropna()

# Convert categorical columns to numerical values using one-hot encoding
df = pd.get_dummies(df, drop_first=True)

# Check the correlation with the target variable (assuming 'Sales' is the target)
corr_matrix = df.corr()
print(corr_matrix['Sales'].sort_values(ascending=False))

# Print the available columns after get_dummies to see the correct names
print(df.columns)
```

```
# Select features with a high correlation to 'Sales'
# Make sure to use the updated column names from the previous print statement
features = df[['Row ID']]  # Replace with actual feature names, 'Order ID' and 'Customer ID' are no longer available
target = df['Sales']

# Split the data into training and testing sets
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.2, random_state=42)
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Initialize and train the model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'R-squared: {r2}')
```

```
# Example of making future predictions
new_data = pd.DataFrame({
    'Row ID':
[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]  #
Replace with actual values
})

# Predict future sales
future_sales = model.predict(new_data)
print(f'Predicted future sales: {future_sales}')
```

OUTPUT FOR THE PROBLEM STATEMENTS

```
Row ID          0
Order ID        0
Order Date      0
Ship Date       0
Ship Mode       0
Customer ID     0
Customer Name   0
Segment         0
Country         0
City            0
State           0
Postal Code     0
```

```
Region          0
Product ID      0
Category        0
Sub-Category    0
Product Name    0
Sales           0
Quantity        0
Discount        0
Profit          0
dtype: int64
Sales       1.000000
Quantity    0.437464
Product Name_HON 5400 Series Task Chairs for Big and Tall    0.291594
Product ID_FUR-CH-10002024                                   0.291594
Product ID_FUR-BO-10004834                                   0.267890
                                                               ...
Product ID_FUR-FU-10004270       -0.056008
Product Name_Staple-based wall hangings    -0.056054
State_Illinois       -0.059166
City_Chicago       -0.060084
Sub-Category_Furnishings       -0.457835
Name: Sales, Length: 6209, dtype: float64
Index(['Row ID', 'Postal Code', 'Sales', 'Quantity', 'Discount', 'Profit',
       'Order ID_CA-2014-100678', 'Order ID_CA-2014-100706',
       'Order ID_CA-2014-100916', 'Order ID_CA-2014-101462',
       ...
       'Product Name_Tensor Brushed Steel Torchiere Floor Lamp',
       'Product Name_Tensor Computer Mounted Lamp',
```

'Product Name_Tensor Track Tree Floor Lamp',
'Product Name_Ultra Commercial Grade Dual Valve Door Closer',
'Product Name_Ultra Door Kickplate, 8"H x 34"W',
'Product Name_Ultra Door Pull Handle',
'Product Name_Ultra Door Push Plate',
'Product Name_Westinghouse Clip-On Gooseneck Lamps',
'Product Name_Westinghouse Floor Lamp with Metal Mesh Shade, Black',
'Product Name_Westinghouse Mesh Shade Clip-On Gooseneck Lamp, Black'],
dtype='object', length=6209)
Mean Squared Error: 306563.5646297518
R-squared: -0.00046797502100925925
Predicted future sales: [355.89500454 355.89411216 355.89321978 355.89232739 355.89143501
355.89054262 355.88965024 355.88875785 355.88786547 355.88697309
355.8860807 355.88518832 355.88429593 355.88340355 355.88251116
355.88161878 355.88072639 355.87983401 355.87894163 355.87804924
355.87715686 355.87626447 355.87537209 355.8744797 355.87358732
355.87269493]