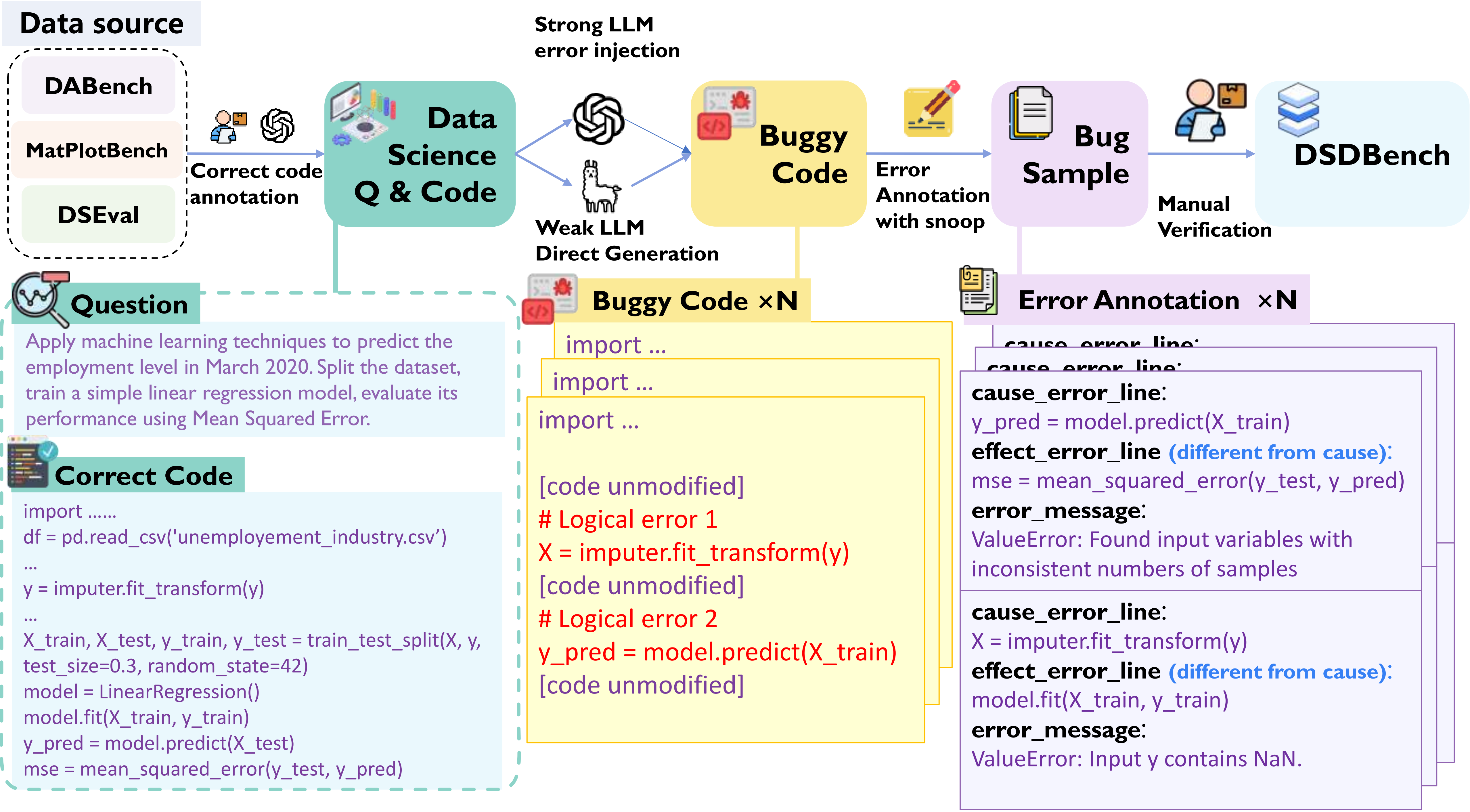

Question

Apply machine learning techniques to predict the employment level in March 2020. Split the dataset, train a simple linear regression model, evaluate its performance using Mean Squared Error.

Correct Code

```
import ......
df = pd.read_csv('unemployement_industry.csv’)
...
y = imputer.fit_transform(y)
...
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
```

Buggy Code ×N

```
import ...
```

```
import ...
```

```
import ...

[code unmodified]
# Logical error 1
X = imputer.fit_transform(y)
[code unmodified]
# Logical error 2
y_pred = model.predict(X_train)
[code unmodified]
```

Error Annotation ×N

cause_error_line:

cause_error_line:

cause_error_line:
y_pred = model.predict(X_train)
effect_error_line **(different from cause)**:
mse = mean_squared_error(y_test, y_pred)
error_message:
ValueError: Found input variables with inconsistent numbers of samples

cause_error_line:
X = imputer.fit_transform(y)
effect_error_line **(different from cause)**:
model.fit(X_train, y_train)
error_message:
ValueError: Input y contains NaN.