“SUMMARY FOR NEXT HIKE PROJECT 6”

### Project Overview

Objective: Predict sales using historical sales data and store information.

### Data Overview

1. Training Data (`train\_p8.csv`): Contains historical sales data including sales amounts, store information, promotions, and holidays.

2. Testing Data (`test.\_p8csv.csv`): Used for making predictions. Includes similar features as the training data.

3. Store Data (`store\_p8.csv`): Contains details about different stores.

4. Submission Data (`sample\_submission\_p8.csv`): Provides a format for submitting predictions.

### Exploratory Data Analysis (EDA)

1. Missing Values: Handled by forward-filling.

2. Duplicates: Checked and found none.

3. Visualizations:

- Sales Distribution: Plotted the distribution of sales.

- Sales by Store: Analyzed how sales vary by store type.

- Promo Impact: Examined sales during promotions versus non-promotional periods.

- Sales vs. Customers: Looked at the relationship between sales and the number of customers.

- Sales During Holidays: Analyzed how sales are affected by different holidays.

### Feature Engineering

1. Date Features: Added features for day of the week and weekend indicators.

2. Scaling: Normalized sales and customer numbers.

### Modeling

1. Random Forest Regressor:

- Pipeline: Preprocessing with imputation and encoding, followed by Random Forest regression.

- Evaluation: Mean Squared Error (MSE) was used to evaluate the model on training data.

2. Deep Learning with LSTM:

- Data Preparation: Created sequences of sales data for LSTM training.

- LSTM Model: Built and trained an LSTM model to predict sales based on sequences of past sales.

### Results

- The Random Forest model's performance was evaluated using MSE.

- The LSTM model was trained and used for predicting future sales based on historical patterns.

### Next Steps

- Tune the models further for better accuracy.

- Make predictions on the test data and prepare the final submission.

This summary covers the key steps and findings in your project, providing a clear overview of your data analysis, feature engineering, and modeling approach...