EXPERIMENT REPORT

Name	Monali Patil
Project Name	Retail-Store-Revenue-Prediction-Api-Service
Deliverables	Model name: Phophet algorithm for Revenue Forecasting Notebook name: MachineLearning_ExploratoryDataAnalysis(EDA).ipynb MachineLearning_ForecastingModel_Phophet.ipynb Project Repo: https://github.com/monalippatil/MachineLearning-Retail-Store-Revenue-Prediction-Api-Service.git

1. EXPERIMENT BACKGROUND	
Provide information about the problem/project such as the scope, the overall objective, expectations. Lay down the goal of this experiment and what are the insights, answers you want to gain or level of performance you are expecting to reach.	
1.a. Business Objective	Explain clearly what is the goal of this project for the business. How will the results be used? What will be the impact of accurate or incorrect results? Goal: The project involves working with an American retail chain operating across California (CA), Texas (TX), and Wisconsin (WI) having 10 stores to develop two distinct machine learning predictive and forecasting models to assist in business decisions related to sales revenue. Application: A predictive model would be utilized to accurately predict sales revenue for individual items in specific stores for any given date crucial for managing inventory levels, optimizing pricing strategies, and ensuring product availability. The forecasting model would be employed to forecast the total sales revenue across all stores and items for the next 7 days allowing for better resource allocation, staffing decisions, and financial planning. Impact: The accuracy of these predictive and forecasting models is crucial for the retailer's operational efficiency, profitability, and customer satisfaction. Correct predictions and forecasts empower the retailer to make informed decisions that positively impact its business revenue growth, while inaccurate results can lead to financial losses with unsold inventory by overstocking, stockouts by understocking, and inadequate staffing, leading to poor customer service and lost revenue.

1.b. Hypothesis	**Present the hypothesis you want to test, the question you want to answer or the insight you are seeking. Explain the reasons why you think it is worthwhile considering it,** Hypothesis: Can the Prophet time series algorithm improve the accuracy and interpretability of revenue forecasts for the retail business? Rationale: Prophet is designed to capture various temporal patterns in time series data, including seasonality, holidays, and trends. By leveraging these patterns, the aim is to improve the accuracy of revenue forecast ultimately benefiting inventory management, pricing strategies, and overall business profitability in the retail sector, while the sought insight pertains to the potential benefits and actionable information it can provide to enhance retail operations. Moreover, the hypothesis aligns with a prevalent objective within the retail industry, which involves harnessing data-driven insights to make well-informed decisions that drive business success.
1.c. Experiment Objective	**Detail what will be the expected outcome of the experiment. If possible, estimate the goal you are expecting. List the possible scenarios resulting from this experiment.** The expected outcome of the Prophet revenue forecasting experiment is to assess whether the algorithm can provide more accurate and interpretable revenue forecasts for the retail business. The retailer's expected outcome of this experiment would be development of a machine learning predictive model utilizing regressor algorithm that can accurately predict the sales revenue for individual items in specific stores for any given date. The goal is to ensure that the forecasting model provides accurate sales revenue for the upcoming 7 days, commencing from the provided input date. Possible scenarios resulting from this experiment include the following situations: * Improved Forecast Accuracy: The Prophet algorithm demonstrates significantly improved forecast accuracy, with Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) measures substantially lower than those obtained with previous methods. This would be a highly successful outcome, indicating the algorithm's suitability for revenue forecasting. * Interpretable Patterns: Expect Prophet algorithm to provide interpretable patterns and insights related to revenue trends, seasonality, and special events (holidays and promotions). These insights could help the retail business understand the underlying factors influencing revenue fluctuations. * Failure: Prophet does not offer substantial improvements in forecast accuracy or interpretability. This could suggest that the algorithm is not well-suited to the retail business's revenue forecasting needs.

2. EXPERIMENT DETAILS

Elaborate on the approach taken for this experiment. List the different steps/techniques used and explain the rationale for choosing them.

2.a. Data Preparation

Describe the steps taken for preparing the data (if any). Explain the rationale why you had to perform these steps. List also the steps you decided to not execute and the reasoning behind it. Highlight any step that may potentially be important for future experiments

The retail data underwent exploration, cleaning, and preparation as part of the subsequent steps required for the utilization of the regression algorithm.

• Data Understanding

1] Loading Data: Imported below given data in separate CSV files into the panda's data frames to use and create a predictive model.

- Training Data
- Evaluation Data
- Calendar
- Events
- Items Price per Week

2] Exploring Data: Examined and studied information using different panda functions to comprehend and uncover patterns, aiming to identify prospective features for revenue prediction. The below tasks are performed to ensure the quality of the data to be utilized by the model analysis.

- Handing missing/null values.
- Eliminating identifiers.
- Combining the individual datasets into a single dataset
- Computing the total revenue based on the items sold and the weekly price of an item
- Processing the ‘date’ feature to derive additional date related information
- Selecting appropriate features
- Identifying and eliminating identifier
- Splitting data into different datasets
- Handling and imputing missing values

df.head(): Checking initial records of the dataset.
df.shape(): Verifying the dimension of the dataset.
df.columns: Identifying attributes name.
df.info(): Assessing the summary information of the attributes of the dataset.
df.describe(include=’all’): Examining statistical summary information for all variables of the dataset across different data types.
df.isnull().sum(): Examining whether there are any null values in the dataset.

• Data Preparation

3] Reshaping the training dataset and renaming the attributes

	id	item_id	dept_id	cat_id	store_id	state_id	d_1	d_2	d_3	d_4	...	d_1532	d_1533	d_1534	d_1535	d_1536	d_1537	d_1538
0	HOBBIES_1_001_CA_1_evaluation	HOBBIES_1_001	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	1	1	1	0	1	0	1
1	HOBBIES_1_002_CA_1_evaluation	HOBBIES_1_002	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	0	0	0	0	0	0	0
2	HOBBIES_1_003_CA_1_evaluation	HOBBIES_1_003	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	0	0	1	0	0	0	0
3	HOBBIES_1_004_CA_1_evaluation	HOBBIES_1_004	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	8	2	0	8	2	3	1
4	HOBBIES_1_005_CA_1_evaluation	HOBBIES_1_005	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	2	0	1	3	2	1	1

* The above training dataset had columns named from d_1 to d_1541 representing the sold item volume for each day for distinct items from 10 stores across 3 separate states. So, melt functionality was utilized to reshape the data to improve attribute organization and for more convenient and suitable data analysis and modelling task.
Additionally, the 'd_*' column was relabeled as 'dayofsale,' and the 'sold item numbers' were renamed as 'volume.

4] Combining the individual datasets into a single dataset

id	item_id	dept_id	cat_id	store_id	state_id	dayofsale	volume	date	wm_yr_wk	d	event_name	event_type	sell_price
BBIES_2_050_CA_4_evaluation	HOBBIES_2_050	HOBBIES_2	HOBBIES	CA_4	CA	d_741	0	2013-02-07	11302	d_741	NaN	NaN	1.97
OODS_3_782_WI_2_evaluation	FOODS_3_782	FOODS_3	FOODS	WI_2	WI	d_114	0	2011-05-22	11117	d_114	NaN	NaN	2.48
HOLD_2_505_CA_1_evaluation	HOUSEHOLD_2_505	HOUSEHOLD_2	HOUSEHOLD	CA_1	CA	d_686	1	2012-12-14	11246	d_686	NaN	NaN	4.97

* Combining data from various sources into a single data frame is a fundamental step in data preprocessing for modelling. It ensures data consistency, enables feature engineering, and simplifies the overall analysis process, ultimately contributing to better decision-making and model performance.

5] Feature Engineering.

Please refer to the 2.b. Feature Engineering section.

6] Selecting appropriate features

Features selected: 'item_id', 'store_id', 'date', 'event_name', 'event_type', 'revenue'

* To build a machine learning predictive model, above specific features were chosen due to the presence of duplicate columns after aggregating the datasets. This duplication resulted in the large dataset leading to memory related issues. For instance, ‘store_id’ field values (CA_1, WI_2, TX_9) already incorporates the ‘state_id’ attribute values (CA, WI, TX). Therefore, it's logical to choose pertinent features that have generalized unique values and avoid duplicates.

* Additionally, to address the memory challenges associated with analyzing the extensive dataset, the aforementioned features were chosen based on their relevance from a business perspective.

7] Identifying and eliminating identifiers

* The 'id' attributes serve as unique identifiers for individual items sold within a specific category, store, and state on a particular date. As a result, during the feature selection process, this identifier was omitted.

* Including it in the analysis could potentially result in overfitting, where the model becomes too focused to these specific values, rather than capturing the underlying, generalized patterns present in the retailer's data.

8] Splitting data into different datasets

To split the dataset into training and validation below strategy was employed.

- Training data: From 2011-01-01 to 2014-12-31
- Validation data: from 2015-01-01 to 2016-12-31

* The retail data covers multiple years, and by splitting it into training (historical) and validation (future) sets, would maintain the temporal consistency of the dataset. This ensures that the model is trained on past data and tested on more recent data, mimicking a real-world scenario.

* Furthermore, this approach aided in handling the relatively compact dataset, mitigating memory-related challenges.

9] Handling and imputing missing values

After combining the datasets, missing values were identified in the attributes.

Attribute Name	Total missing values
event_name	43143350
event_type	43143350
revenue	12291876

* To handle missing values in the 'event_name' and 'event_type' features, 'None' was used as a replacement for the missing values.

* For address missing values in the 'revenue' target feature, 0 was used for imputation. This signifies days with no items sold and, consequently, zero income.

10] Aggregating the revenues based on the date across all items, stores, and states.

	date	revenue
0	2011-01-29	81650.61
1	2011-01-30	78970.57
2	2011-01-31	57706.91
3	2011-02-01	60761.20
4	2011-02-02	46959.95

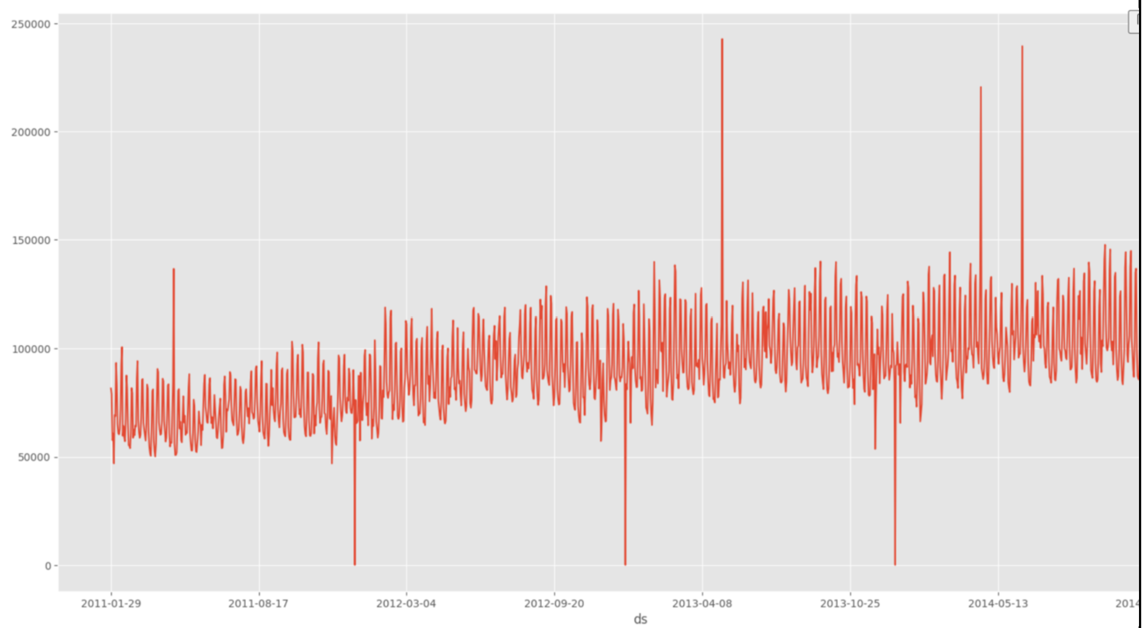
To construct the revenue forecasting model using the Prophet time-series model, aggregated the revenue figures based on the date, encompassing all items, stores, and states.

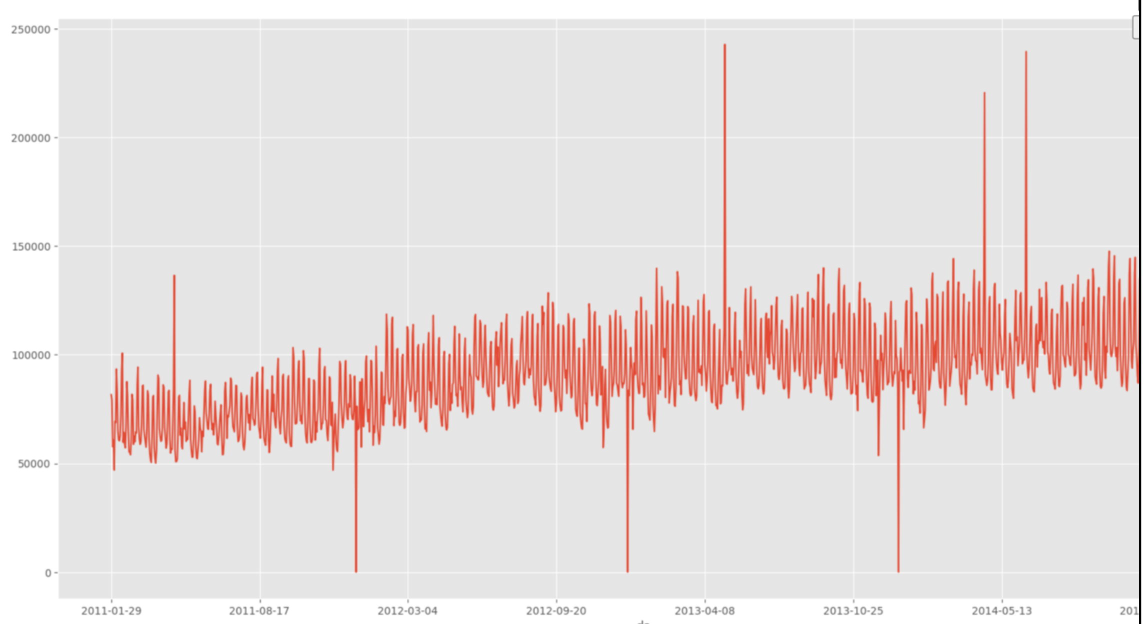
The following measures are crucial and may hold significance for any future regression experiments.

	1. When working with date-related datasets and business scenarios, it's vital to split the data into training and validation sets based on the date. This practice helps prevent data leakage, where information from the validation period unintentionally influences the training process. This ensures a fair assessment of the model's predictive performance. 2. To prevent the model from becoming too specialized on particular data points and to enable it to learn generalized patterns, it's essential to eliminate any identifying attributes. 3. Carefully evaluating and handling missing values is crucial to build machine learning models and avoid introducing biases into the model.
2.b. Feature Engineering	Describe the steps taken for generating features (if any). Explain the rationale why you had to perform these steps. List also the feature you decided to remove and the reasoning behind it. Highlight any feature that may potentially be important for future experiments - Computing the total revenue based on the items sold and the weekly price of an item * Calculating the overall revenue by considering the quantity of items sold and their corresponding weekly prices was essential for developing the machine learning model to predict target variable revenue. - Processing the 'date' feature to derive additional date related information * To enhance the dataset's date related information by extracting additional date related features from the 'date' attribute such as day of a week, month, year, and week of the year for model to learn generalized features. * Therefore, set the datetime type to the date feature and utilized the datetime functionality to extract the day of the week, month, year, and week of the year information.
2.c. Modelling	Describe the model(s) trained for this experiment and why you choose them. List the hyperparameter tuned and the values tested and also the rationale why you choose them. List also the models you decided to not train and the reasoning behind it. Highlight any model or hyperparameter that may potentially be important for future experiments Regression Algorithm: The target variable 'revenue' contains continuous values that represent sales income. Since our goal is to predict these continuous incomes, using a regression algorithm is the appropriate choice, aligning with the objective of learning and prediction. Algorithm Name / Rationale

1. Phophet time series algorithm	* The Prophet time series model for revenue forecasting has the ability to handle time-dependent data effectively. * Prophet is specifically designed for forecasting tasks and can capture the various patterns and seasonality often present in time series data, which is common in revenue forecasting. * Additionally, it provides the flexibility to incorporate external factors, such as holidays and events, which can significantly impact sales. * Moreover, Prophet is user-friendly and requires minimal tuning, making it a suitable choice for both beginners and experts. Overall, its predictive accuracy and ease of use make it a strong candidate for revenue forecasting in this learning task.

Employing Holiday Construct of the Phophet algorithm.





	* The charts above demonstrate periodic spikes in retail revenue in both directions, with noticeable increases and declines occurring during specific months of the year, following a recurring pattern at certain intervals. * While these spikes could be attributed to various factors, we can hypothesize that they may be related to long weekends, holidays, or significant promotions that the company runs approximately every quarter, potentially impacting the sales revenue. * To leverage this information for more accurate forecasting during those months, utilising Prophet's holiday feature which is a Pandas dataframe containing the holidays and their respective dates. Due to time constraints, there was no chance to explore time series forecasting models like SARIMA.

3. EXPERIMENT RESULTS

Analyse in detail the results achieved from this experiment from a technical and business perspective. Not only report performance metrics results but also any interpretation on model features, incorrect results, risks identified.

3.a. Technical Performance

Score of the relevant performance metric(s). Provide analysis on the main underperforming cases/observations and potential root causes.

The table below informs the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) performance scores of the Phophet algorithm on both training and validation datasets.

Algorithm	Datasets	MAPE Score	RMSE Score
	Baseline Performance	898.0839	21278.1704
Experiment 1: Phophet algorithm	Training Dataset	821.0444	10442.7834
	Validation Dataset	7.0549	9861.6278

The baseline model's MAPE score of 898.0839 indicates that, on average, it predicted sales revenue values with an error of approximately 898%.

The RMSE score of 21278.1704 represents the root mean square error, which measures the square root of the average squared differences between the predicted and actual revenue values. This score is exceptionally elevated, suggesting a significant spread of errors in the baseline model's revenue forecasts.

Experiment 1: Phophet algorithm

* The Prophet algorithm's, MAPE score on the training dataset reduced to 821.0444, indicating that the Prophet algorithm reduced the average prediction error by a significant margin.

* The RMSE score in the training dataset decreased to 10442.7834, signifying a substantial reduction in prediction errors' dispersion.

* The validation dataset showed remarkable results, with an exceptionally low MAPE score of 7.0549. This indicates that, on average, the Prophet algorithm's predictions deviated by only about 7% from the actual sales revenue values, demonstrating high accuracy.

* The RMSE score in the validation dataset further decreased to 9861.6278, highlighting the effectiveness of the Prophet algorithm in reducing errors in revenue predictions.

3.b. Business Impact	Interpret the results of the experiments related to the business objective set earlier. Estimate the impacts of the incorrect results for the business (some results may have more impact compared to others) * The experiments using the Prophet algorithm showed significant improvements in terms of both MAPE and RMSE scores. The MAPE score for the validation dataset, in particular, dropped significantly, indicating that the Prophet model was able to provide more accurate revenue forecasts. This can result in improved inventory management, better pricing decisions, and ultimately higher profitability for the business. * The model's ability to capture time series patterns and seasonality led to significantly improved forecasts. These results suggest that the Prophet algorithm is a promising choice for revenue forecasting in this context, with the potential for further fine-tuning and optimization. * Additionally, the results indicate that employing the Prophet time series forecasting model can significantly benefit the business by providing more accurate revenue predictions, which can inform better decision-making processes.
3.c. Encountered Issues	List all the issues you faced during the experiments (solved and unsolved). Present solutions or workarounds for overcoming them. Highlight also the issues that may have to be dealt with in future experiments. * Memory constraints was a challenge, particularly during EDA tasks involving the large dataset. Nonetheless, this issue was successfully addressed by splitting the dataset by date and selecting pertinent features. * During the pipeline construction, could not employ one-hot encoding for categorical feature transformation with a large number of unique values in categorical features. This approach significantly increased the dimensionality of the features, leading to excessively long processing times before encountering system crashes. Therefore, ordinal encoder was utilised.

---

4. FUTURE EXPERIMENT	
Reflect on the experiment and highlight the key information/insights you gained from it that are valuable for the overall project objectives from a technical and business perspective.	
4.a. Key Learning	Reflect on the outcome of the experiment and list the new insights you gained from it. Provide rationale for pursuing more experimentation with the current approach or call out if you think it is a dead end. * While the Prophet model showed better performance, it is important to note that the experiments were conducted without hyperparameter tuning due to time constraints. Further experimentation with hyperparameter optimization could lead to even better results. * The quality of data, as well as feature engineering, play crucial roles in forecasting accuracy. Also, it is important to perform additional evaluations, such as cross-validation and robustness testing, to ensure that the model performs well under various conditions and doesn't overfit the training data.

	* Additionally, exploring other time series forecasting models like SARIMA is a valuable avenue for future research as it is essential to consider a variety of models to identify the one that best suits the business's specific needs.
4.b. Suggestions / Recommendations	Given the results achieved and the overall objective of the project, list the potential next steps and experiments. For each of them assess the expected uplift or gains and rank them accordingly. If the experiment achieved the required outcome for the business, recommend the steps to deploy this solution into production. Will conduct below activities. * Perform hyperparameter tuning for the Prophet model, optimizing parameters related to seasonality, holidays, growth, and changepoints. Would employ, Hyperopt, Grid Search, or Random Search for this purpose. * Implement cross-validation techniques like time series cross-validation to assess the model's performance across different time periods, ensuring it generalizes well to unseen data * Explore ensemble methods, such as combining the Prophet model with other time series models namely SARIMA.