Capstone Proposal 3:

**[E-COMMERCE RETAIL SALES ANALYSIS]**

**Business Understanding:**

**Problem Statement:** This project aims to predict daily amounts of sold products.

**Relevance:** As storage area may be expensive and fast delivery on time is important to prevail over the competition we like to help the retailer by predicting daily amounts of sold products to identify peak periods per type of product and overall.

**Industry Domain:** This applies to all retail gift shops in the UK.

**Target Audience:** The primary audience will be gift retail store owners, investors, and suppliers.

**Impact:** By accurately predicting sales at different times of the year, this could advise gift stores on how much stock the need to have at a given time for maximum sales without incurring an extra cost on storage of idle stock.

**Research:** Exploration of the most popular products and fast moving product for a retail gift store.

**Data Understanding:**

**Data Collection:** Data to be obtained from UCI website [The UCI Machine Learning Repository](http://archive.ics.uci.edu/ml/index.php).

**Context:**  This is a dataset containing actual transactions from 2010 and 2011 from a retail store In the UK.

**Comparative Analysis:** Various predictive models and analytical methods will be explored to identify gaps and opportunities for innovation.

**Data Preparation:**

**Data Types:** We anticipate a mix of continuous and categorical data.

**Preprocessing Steps:** Anticipated steps include data cleaning, dealing with missing data, feature engineering, normalization and encoding categorical variables.

**Challenges:** inconsistencies in the data set.

**Data Volume:** Aiming for a large dataset (thousands) to train and test our model.

**Visualization:** sales trend, income trend, product popularity among others.

**Modeling:**

**Techniques:** Time series analysis for sales predictions.

**Target Variable:** The target variables will be daily sales and revenue.

**Baseline Model:** A linear regression model.

**Problem Type:** Daily sales predictions.

**Evaluation:**

**Success Metrics:** MSE and RMSE metrics

**MVP:** A model that accurately predicts sales peak seasons based on historical data.

**Deployment**

**Reporting:** Results will be shared through an interactive dashboard that visualizes sales vs. revenue.

**Deployment Plan:** The goal is to help retail stores to know which product to stock at what time of the year.

**Tools/Methodologies**

**Python Libraries:** UCI data for data source, Pandas for data manipulation, NumPy for numerical operations, Matplotlib/Seaborn for visualization, Scikit-learn for modeling.

**Algorithms:** Linear Regression, ARIMA for time series prediction.

**Environment:** Analysis will be conducted locally with potential use of cloud-based resources for increased computational power.

**Data Storage:** Initially on local machine, considering cloud storage (AWS S3) for larger datasets or if web deployment is pursued.