**Week5\_Introduction of project\_shahe**

**1. Introduction**   
  
**1.1 Project Overview**

This project aims to develop a web-based software solution for **Shirley Road Records**, a Durham-based music business, to maximize projected ticket sales for concerts. The solution will automate the data collection process from multiple concert venues, perform ticket sales analysis using regression models, and provide users with a dashboard to manage, view, and analyze the collected data.

Currently, Shirley Road Records manually collects data from concerts they book and other publicly available sources. The company performs regression analysis in R to rank artists based on ticket sales potential. To streamline this process, the project will expand an existing Python web scraper and build an interface that allows the user to manage and analyze data more efficiently.

**1.2 Project Objectives**

1. **Automate Data Collection**: Develop a robust web scraper that collects concert data (e.g., band/artist name, venue, ticket price, and day of the week) from multiple venue websites.
2. **Data Management**: Create a system that organizes, stores, and processes scraped data for use in analysis. This could involve exporting the data to a database or Excel sheet.
3. **Sales Prediction and Analysis**: Integrate R language regression analysis into the system to predict which concerts are likely to have high ticket sales based on historical and newly scraped data.
4. **User Interface**: Design a user-friendly interface that allows users to:
   1. View, manage, and edit collected concert data.
   2. Run analysis tasks to predict ticket sales for different bands and venues.
   3. Display the analysis results in an intuitive manner.
5. **Seamless Integration**: Ensure smooth interaction between the front-end (UI), backend (web scraper and data processor), and analysis tools.
6. **Maximize Ticket Sales**: Provide actionable insights for concert booking decisions by identifying bands and venues that perform well in terms of ticket sales.

### **1.3 Project Scope**

The scope of this project includes:

1. **Development of Multi-Venue Web Scrapers**:
   1. Extend the existing web scraper to collect data from multiple concert venues.
   2. Ensure flexibility to easily add or modify scraping for new venues as needed.
   3. Data fields to be collected include band/artist, venue, ticket price, and date.
2. **Backend Data Processing**:
   1. Data normalization and formatting for easy insertion into an Excel sheet or database.
   2. Implement a system to process and prepare scraped data for analysis.
3. **R Integration for Analysis**:
   1. Use R for regression analysis on the data to predict ticket sales potential.
   2. Create a workflow where users can trigger R scripts to run analysis on selected data.
4. **User Interface Development**:
   1. Build a front-end dashboard where users can view, manage, and analyze data.
   2. Display sales projections and analysis results in a clear, intuitive format (e.g., charts, graphs, ranked lists).
5. **System Testing and Validation**:
   1. Validate that the web scraper collects accurate and complete data.
   2. Ensure the system integrates correctly with the R analysis tools.
   3. Perform performance testing to ensure the UI and data analysis processes are smooth and user-friendly.

#### **Excluded Features**

1. **User Login and Access Management**: No need to develop complex user management functionality.
2. **Advanced Recommendation System**: Only basic filtering and sorting of data are supported, without complex AI-based recommendation algorithms.
3. **Mobile Platform Support**: The project will focus on web development, without mobile platform support.
4. **Internationalization Support**: The project will only support English and will not include multi-language versions.

#### **Project Constraints and Assumptions**

1. **Time Constraints**: The project will be completed within four sprint cycles.
2. **Technical Requirements**:
   1. Ensure that the scraper tool can bypass anti-scraping mechanisms to reliably collect data.
   2. The database must support efficient data querying and filtering for sales prediction analysis.
3. **Dependencies**:
   1. The project depends on third-party websites (venue websites) for data. Changes in these sources may affect project progress.

This software will streamline decision-making processes, allowing Shirley Road Records to optimize concert bookings and maximize ticket sales through a data-driven approach.