Midterm Exam			
Topic:	**Time Series Analysis and Forecasting with ARIMA Model**	Week No.	10
Course Code:	**CSST104**	Term:	2nd Semester
Course Title:	**Machine Learning**	Academic Year:	2023-2024
Student Name		Section	
Due date		Points	

Forecasting with ARIMA Model

Objective:
To apply ARIMA modeling techniques to forecast the number of app updates in the Google Play Store for the next 12 months based on historical data.

Dataset:

Google Play Store Dataset
- **Description:** Contains information about the google play store.
- **Use Case:** Ideal for time series analysis to predict continuous outcomes like the downloads based on various features (e.g., ratings, reviews, installs).

1. Data Cleaning and Preparation
- **Task**: Clean the dataset by handling missing values, removing duplicates, and converting fields like 'Reviews', 'Size', 'Installs', and 'Price' to a numerical format for analysis.
- **Skills Assessed**: Data cleaning, data type conversion.

2. Exploratory Data Analysis (EDA)
- **Task**: Perform EDA to uncover trends in the dataset. This may include analyzing the distribution of app ratings, understanding the proportion of free vs. paid apps, and identifying which categories have the most apps.
- **Skills Assessed**: Data visualization, descriptive statistics, pandas proficiency.

3. Category Analysis
- **Task**: Analyze the app categories to find out which category has the highest average rating, the most reviews, and the most apps. Additionally, investigate the relationship between category and app size or price.
- **Skills Assessed**: Groupby operations, aggregation, data visualization.

4. Rating Prediction Model
- **Task**: Build a model to predict the rating of an app based on its features (e.g., Reviews, Size, Installs, Type, Price, Content Rating). Split the data into training and testing sets, select relevant features, and evaluate the model's performance.
- **Skills Assessed**: Machine learning model building, feature selection, model evaluation.

5. Trend Analysis
- **Task**: Analyze the trends over time in the app market. Investigate how the characteristics of the top apps (e.g., rating, reviews, installs) have evolved. Consider the impact of Android versions on app ratings or installs.
- **Skills Assessed**: Time series analysis, trend visualization.

6. Impact of Reviews on Ratings
- **Task**: Investigate the correlation between the number of reviews an app receives and its rating. Determine if more reviews are indicative of higher ratings.
- **Skills Assessed**: Correlation analysis, scatter plot visualization.

These tasks are designed to assess a variety of data science skills, from basic data manipulation to more complex analyses and model building. Feel free to adapt or expand upon these tasks based on the specific objectives of your assessment.

This assessment task evaluates proficiency in time series analysis, specifically the application of ARIMA models, and the ability to interpret and communicate findings from complex analyses.

Submission Instruction:
- Share the Google Collab Activity to markbernardino@lspu.edu.ph
- Filename Format: **2A-BERNARDINO-MIDTERM**

Inability to follow this instruction will be deducted 5 points each for filename format and late submission per day. Also, cheating and plagiarism will be penalized.