

Step 1: Read the Scenario

SCENARIO: As a data analyst at TechGear Inc., a company specializing in electronic gadgets and accessories, your task is to analyze historical sales data, build predictive models, and use prescriptive analytical methods to provide actionable insights for improving decision-making. The company has been experiencing fluctuating sales and aims to optimize its marketing strategies and production processes to maximize profits and enhance customer satisfaction. Your analysis will help TechGear Inc. understand the factors influencing its sales, forecast future sales trends, assess financial risks associated with different business scenarios, and determine the optimal allocation of its marketing budget and production resources. Ultimately, your work will enable the company to make data-driven decisions, enhancing its sales and marketing strategies, and leading to improved profitability and customer satisfaction.

Step 2: Look Over the Data

- Questions 1-5 (Linear Regression) and 7 (Machine Learning): Use the data in the techgear_sales_data.xlsx Excel file, which can be found at the following GitHub link:
https://raw.githubusercontent.com/sophiaAcademics/BDA_Excel/main/StudentData/Assessments/Unit6/TechGearTouchstone/techgear_sales_data.xlsx
- Question 6 (Forecasting): Use the data in the techgear_sales_data_monthly.xlsx Excel file, which is available at this GitHub link:
https://raw.githubusercontent.com/sophiaAcademics/BDA_Excel/main/StudentData/Assessments/Unit6/TechGearTouchstone/techgear_sales_data_monthly.xlsx
This file contains the same data as techgear_sales_data.xlsx, but the last row only includes a date with missing values for all other columns. These missing values are intended for you to apply forecasting methods for the upcoming time period.
- Questions 8 and 9: Since Question 8 focuses on Monte Carlo simulations and Question 9 focuses on linear programming, all necessary data is provided in the problem statement.

This dataset contains monthly sales and advertising spend data for TechGear from January 2020 to December 2024. It includes the following columns:

Column Name	Description	Unit/Format
Date	The month and year for each data entry	MM/DD/YYYY
Sales	The total sales generated in that month	Number of Sales

Ad_Spend_Facebook	The amount of money spent on Facebook advertising in that month	Dollars
Ad_Spend_Instagram	The amount of money spent on Instagram advertising in that month	Dollars
Discount_Rate	The discount rate applied to sales in that month	Percentage

A snapshot of the first few rows of the dataset is provided below:

	A	B	C	D	E
1	Date	Sales	Ad_Spend_Facebook	Ad_Spend_Instagram	Discount_Rate
2	1/31/2020	4283	340	146	10
3	2/29/2020	4775	284	200	6
4	3/31/2020	5437	290	240	10
5	4/30/2020	4541	269	193	10
6	5/31/2020	3161	218	182	16

Step 3: Read TechGear Inc. Questions

Question 1: Exploring Data Structures and Averages in Advertising Spend and Discounts

Before conducting an analysis, use Python to create a pandas DataFrame named sales from the dataset.

- What key features of the dataset can you summarize, such as the number of rows and columns?
- What is the average amount spent on advertising for each social media platform (Facebook and Instagram)?
- What is the average discount provided to customers?
- What insights can you draw from this summary regarding advertising spend and discount trends?

Question 2: Visualizing Relationships

- How can you visualize the relationships between sales and each advertising spend variable (Facebook and Instagram) as well as discount rates?
- What types of plots (e.g., scatter plots, line plots, or histograms) would be most effective in identifying patterns or correlations between these variables?

- What do these visualizations reveal about the impact of advertising spend and discount rates on sales?

Question 3: Simple Linear Regression

TechGear wants to optimize its marketing strategy.

- How can you develop a simple linear regression model in Python to predict sales based on Facebook ad spend?
- What do the coefficients of the model indicate?
- Specifically, how does the slope describe the relationship between Facebook ad spend and sales?
- What does the R^2 value tell you about how well the model explains the variability in sales?
- How does the regression output from Python support your interpretation of the model's performance?

Question 4: Assessing the Fit of the Simple Linear Regression Model

- How can you evaluate the performance of your simple linear regression model by analyzing residuals?
- What insights do residual plots provide about the model's accuracy?
 - Do they suggest any patterns, heteroscedasticity, or violations of linear regression assumptions?
- How might these findings impact the reliability of the model's predictions?

Question 5: Multiple Linear Regression Model

The simple linear regression model provides insights into Facebook ad spend.

- How can you develop a multiple linear regression model to predict monthly sales using Facebook ad spend, Instagram ad spend, and discount rates?
- How do the coefficients of this model compare to the simple linear regression model? What do they reveal about the combined influence of these factors on sales?
- Which model performs better in predicting sales?
- How can you compare the effectiveness using statistical metrics (such as R^2 and RMSE)?
- Based on this comparison, what recommendations can you provide to TechGear for optimizing its advertising strategy?

Question 6: Forecasting

Using historical sales data, how can you construct:

- A 3-month moving average forecast for January 2025?
- An exponential smoothing forecast with a smoothing parameter of 0.80 for January 2025?

Given TechGear's preference for emphasizing recent sales trends:

- Which forecasting method provides the most reliable prediction for January 2025?
- What key differences exist between the two forecasting methods, and what do they imply for forecasting accuracy?

Based on your analysis, consider:

- What actionable recommendations can you provide to TechGear to improve its marketing strategies and production planning?

Question 7: TechGear needs a reliable model to predict future sales.

- How can you build and compare different predictive models to achieve this?
- How can you develop a multiple linear regression model using 5-fold cross-validation to predict future sales?
- How can you develop a decision tree model using 5-fold cross-validation to predict future sales?
- How do the two models compare in terms of RMSE, and which model should TechGear choose?

TechGear requires a minimum of \$6,500 in sales each month to remain profitable.

- If the best model predicts sales of \$4,200, how can the RMSE value be used to determine the range within which actual sales may fall?
- What are the implications of this for decision-making and risk assessment?

Question 8: Monte Carlo Simulations

TechGear has experienced significant fluctuations in sales, making accurate predictions challenging.

- How can you use Monte Carlo simulations to estimate future sales?
- How can you estimate the average and median monthly sales by running 1,000 simulations?

- What visuals (e.g., histograms or box plots) can you generate to summarize the results?
- If daily sales are assumed to follow a uniform distribution between the minimum and maximum observed sales over the past 60 months, how does this impact the simulation results? You can assume that the value for minimum sales observed over 60 months is 2,299 and the maximum value is 7,702.
- How can you interpret the standard deviation of simulated sales, and what does it reveal about TechGear's sales variability?
- How can TechGear use these insights to improve budgeting, sales forecasting, and operational decision-making?

Question 9: Linear Programing

TechGear wants to optimize its advertising spend across Facebook and Instagram to maximize its monthly sales. They have a fixed advertising budget and need to determine the optimal allocation of this budget to achieve the highest possible sales. The sales generated from advertising on each platform are influenced by the amount spent on that platform.

TechGear has a monthly advertising budget of \$10,000. The estimated sales generated from advertising on Facebook and Instagram are given by the following linear equations:

- Sales from Facebook advertising: $SF = 5F$, where F is the amount spent on Facebook advertising
- Sales from Instagram advertising: $SI = 3I$, where I is the amount spent on Instagram advertising

TechGear has a monthly advertising budget of \$10,000. They must spend at least \$2,000 on Facebook advertising to maintain its presence on the platform. Additionally, they must spend a minimum of \$1,000 and no more than \$7,000 on Instagram advertising due to platform-specific constraints. The amount spent on Instagram advertising should be at least 50% of the amount spent on Facebook advertising to ensure balanced marketing efforts.

- What is the optimal budget allocation for Facebook and Instagram, and what is the maximum sales revenue TechGear can achieve under these conditions?

Step 4: Using the PowerPoint Template, Analyze Data for TechGear Inc.

- Your task is to analyze historical sales data for TechGear Inc. using various analytical techniques.

- You'll apply concepts from linear regression, forecasting, machine learning, and prescriptive analytics.
- The goal is to provide actionable insights to help TechGear make data-driven decisions.
- Include Python code snippets in your slides for data exploration, regression models, forecasting, machine learning, Monte Carlo simulation, and linear programming tasks.
- Your Python code should be accurate and well-documented to demonstrate how each analysis step was performed.
- Your findings will be presented in a PowerPoint presentation, with speaker notes explaining your approach and insights.

Review each question and then follow the directions outlined on each slide to summarize and present your findings for each question.

[TechGearStudentTemplate.pptx](#)

Step 5: Review the Grading Rubric to Ensure All Criteria are Met

Review the rubric to ensure that you understand how you will be evaluated. Also review the requirements to ensure that your Touchstone is complete.

Step 6: Submit Your Touchstone

Submit your completed Touchstone (as a .pptx file) using the blue button at the top of this page.

B. Rubric

Python Analysis (Shown at Key Steps) The inclusion of well-documented, accurate Python code for data exploration, regression models, forecasting, machine learning, Monte Carlo simulation, and linear programming. (5%)	Python code is shown for all major steps, including data exploration, visualization, regression models, forecasting, machine learning, Monte Carlo simulation, and linear programming. Code is well-documented and accurate.	Python code is shown for most key steps. Minor issues with code documentation or accuracy.	Python code is shown for some steps, but critical components are missing or incomplete.	Python code is partially shown but lacks key analyses or is significantly incorrect.	No Python code is provided.
Data Exploration and Summary (Slide 2) Clear summary of data structure, accurate calculation of averages, and key insights from data exploration. Python analysis is included and well-integrated. (10%)	There is a comprehensive summary of data structure with accurate calculation of averages and clear insights from the exploration. Python analysis is included and well-integrated.	Data summary is mostly accurate, with minor errors or missing insights. Python analysis is included.	Basic summary provided, but some key features are missing or inaccurate. Python analysis is incomplete.	Minimal data exploration with several inaccuracies and no significant insights. Python analysis is missing or incorrect.	No data exploration is provided.

Visualizing Relationships (Slide 3) Accurate and clear visualizations showing relationships between sales, ad spend, and discount rate. Proper interpretation of patterns and correlations. (10%)	Clear and accurate visualizations for all specified variables with detailed insights into patterns and correlations. Python-generated plots are used.	Visualizations are mostly accurate and provide useful insights. Minor errors in interpretation or plot generation.	Basic visualizations are provided, but significant patterns or correlations are overlooked. Python plots are incomplete.	Visualizations are unclear or inaccurate with limited analysis. Missing Python plots.	No visualizations are provided.
Simple Linear Regression & Model Fit (Slides 4 & 5) Well-implemented regression model with correct interpretation of coefficients and R^2 value. Assessment of model fit through residual analysis. (10%)	Accurate regression model with clear interpretation of coefficients and R^2 value. Residual plots are well-explained, and the fit is thoroughly assessed. Python output included.	Regression model and assessment are mostly accurate, with minor errors or incomplete explanations.	Basic model output provided, but interpretations and model fit assessments are incomplete or contain errors.	Model is poorly developed, with incorrect interpretations and no reliable assessment of fit.	No regression model or assessment is provided.
Multiple Linear Regression (Slide 6)	Complete and accurate multiple linear regression analysis, with well-explained	Multiple regression analysis is mostly accurate, with minor errors or	Basic multiple regression is provided, but interpretations and	Incomplete or incorrect multiple regression model with minimal explanation.	No multiple regression model is provided.

Complete multiple regression analysis, including variable interpretation and comparison to simple regression. Python output included. (10%)	coefficients and comparison to the simple linear regression model. Python output included.	incomplete comparisons.	comparisons are incomplete or partially inaccurate.		
Forecasting (Slide 7) Implementation of both forecasting methods, clear comparison, and justified selection of the best method based on business needs. (10%)	Both forecasting methods are accurately implemented and compared. The recommendation is well-justified and aligned with TechGear's preferences. Python output included.	Forecasting analysis is mostly accurate, with minor errors or incomplete justification of the chosen method.	Basic forecasting analysis is provided, but one method may be missing, or justification is unclear.	Minimal forecasting analysis with significant errors and no clear recommendation.	No forecasting analysis is provided.
Machine Learning Models (Slide 8) Accurate implementation of multiple regression and decision tree models with RMSE comparison and well-supported model selection. (10%)	Both models are accurately built and compared using RMSE. Clear model recommendation with actionable insights. Python output included.	Machine learning analysis is mostly accurate, with minor errors in the comparison or recommendation.	Basic models are provided, but the comparison and recommendation are incomplete or unclear.	Models are incomplete or contain major errors. Limited or no comparison is provided.	No machine learning analysis is provided.

Monte Carlo Simulations (Slide 9) Simulation correctly executed with proper assumptions, visualizations, and interpretation of results. Actionable insights are provided. (10%)	Simulation is well-executed with clear visualizations and interpretation of results. Actionable insights are provided. Python output included.	Simulation is mostly accurate, with minor errors or incomplete insights.	Basic simulation is provided, but interpretation is incomplete or unclear.	Simulation is incomplete or incorrect with minimal explanation.	No simulation is provided.
Linear Programing (Slide 10) Accurate optimization model that meets constraints and clearly explains the best budget allocation for maximum sales. (10%)	Linear programming solution is accurate and fully meets all constraints. Clear explanation of the optimal budget allocation and maximum achievable sales. Python output included.	Solution is mostly accurate, with minor errors in constraints or explanation.	Basic linear programming solution is provided but contains errors or incomplete explanations.	Incomplete or incorrect solution with minimal explanation.	No linear programming solution is provided.
Presentation Quality & Speaker Notes Well-organized slides with readable formatting and professional layout. Speaker notes	Slides are visually appealing and well-organized, with clear speaker notes that thoroughly explain the analysis and findings.	Slides are mostly clear and organized. Speaker notes are informative but may lack detail.	Basic slides with limited visual appeal. Speaker notes are incomplete or too brief.	Poorly organized slides with missing or unclear speaker notes.	No presentation or speaker notes provided.

effectively explain analysis and insights. (15%)					
--	--	--	--	--	--

C. Requirements

The following requirements must be met for your submission:

- Hand in a .pptx file with slides listed above.
- Use a readable 11- or 12-point font.
- All writing must be appropriate for an academic context. Follow academic writing conventions (correct grammar, spelling, punctuation, and formatting).
- Plagiarism of any kind is strictly prohibited.
- Submission must include your name and the date (included in the template).

This assignment provides a practical experience in business analytics, honing skills essential for data-driven decision-making in business environments. Your analysis and recommendations will help TechGear optimize its operations,