Report.docx

Turnitin

Document Details

Submission ID

trn:oid:::9832:73396280

Submission Date

Dec 8, 2024, 3:35 PM UTC

Download Date

Dec 8, 2024, 3:36 PM UTC

File Name

Report.docx

File Size

631.7 KB

3 Pages

494 Words

3,060 Characters

0% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

1 AI-generated only 0%
Likely AI-generated text from a large-language model.

2 AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

Bike Sharing Analysis and Predictions

Student Name: Abdaal Khan

Student ID: 23073481

GitHub Repository:

Introduction

This report analyzes bike-sharing data to uncover patterns in user behavior and provide insights into rental trends. The dataset includes information about environmental factors, rental counts, and user types. The objective is to apply clustering, regression, and predictive analytics techniques to derive actionable insights. The report is structured as follows:

1. Exploratory data analysis.
2. Clustering analysis using K-Means.
3. Regression analysis to study temperature effects on rentals.
4. Predictive modeling for future rental forecasts.

Data Preprocessing

The dataset underwent preprocessing to prepare for analysis:

- Numerical columns (temp, atemp, hum, windspeed) were normalized.
- Categorical columns were encoded for better usability.
- New features, such as total_users, were engineered by aggregating casual and registered counts.

	holiday	weekday	workingday	weathersit	temp
count	730.000000	730.000000	730.000000	730.000000	730.000000
unique	NaN	NaN	NaN	NaN	NaN
top	NaN	NaN	NaN	NaN	NaN
freq	NaN	NaN	NaN	NaN	NaN
mean	0.028767	2.997260	0.683562	1.394521	20.319259
std	0.167266	2.006161	0.465405	0.544807	7.506729
min	0.000000	0.000000	0.000000	1.000000	2.424346
25%	0.000000	1.000000	0.000000	1.000000	13.811885
50%	0.000000	3.000000	1.000000	1.000000	20.465826
75%	0.000000	5.000000	1.000000	2.000000	26.880615
max	1.000000	6.000000	1.000000	3.000000	35.328347

Exploratory Data Analysis

Categorical Trends

To understand the seasonal distribution of bike rentals, the average number of total users was plotted for each season.

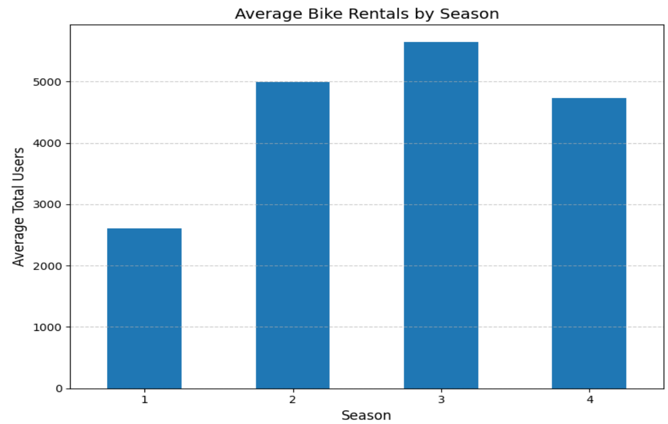

Figure 1 Average Bike Rentals by Season

Key insights:

- Season 3 has the highest rentals, followed by Season 2 and Season 4.
- Season 1 (winter) shows the lowest number of rentals.

Relational Analysis

A scatter plot was created to explore the relationship between temperature and bike rentals.

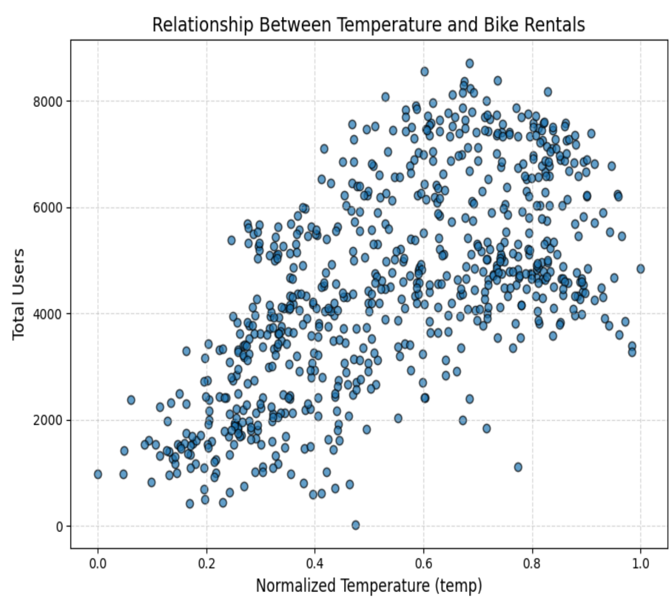

Figure 2 Relationship Between Temperature and Bike Rentals

Key insights:

- A positive correlation exists between temperature and rentals. Warmer weather leads to more bike rentals.

Statistical Analysis

A boxplot was used to analyze rental distributions across different weather conditions.

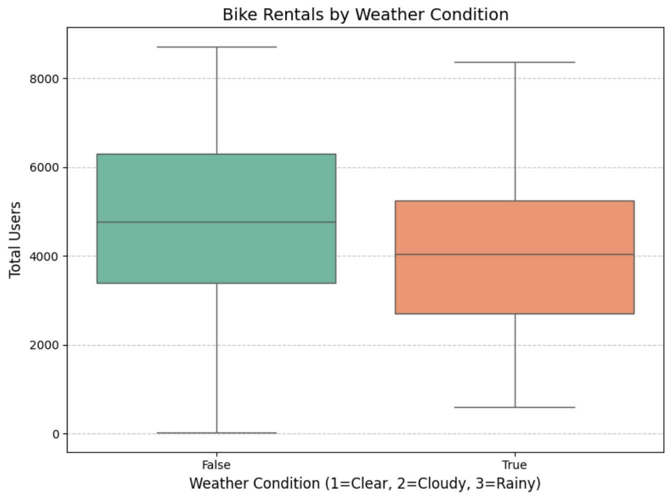

Figure 3 Bike Rentals by Weather Condition

Key insights:

- Clear weather conditions (Weathersit 1) lead to significantly higher rentals compared to cloudy and rainy conditions.

Clustering Analysis

The Elbow Method was employed to determine the optimal number of clusters for segmentation, followed by a visualization of the clustering results.

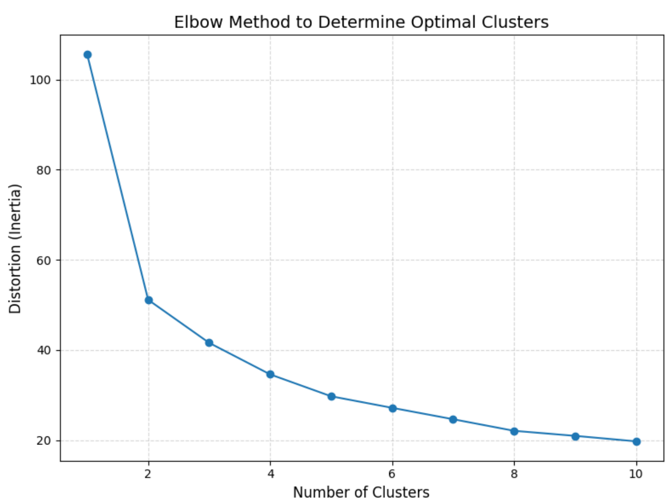

Figure 4 Elbow Method to Determine Optimal Clusters

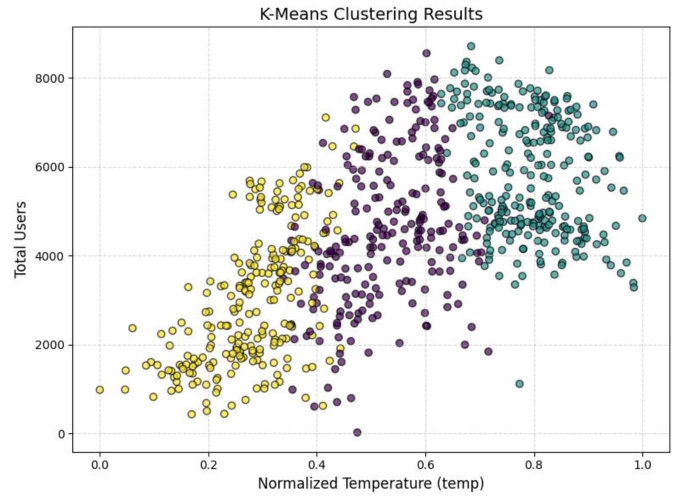

Figure 5 K-Means Clustering

Key insights:

- The Elbow Method suggested 3 clusters as the optimal number.
- Clustering analysis grouped rentals into meaningful clusters based on temperature and other features.

Regression Analysis

A linear regression model was built to study the impact of temperature on bike rentals. The fitted line and model metrics provide insights into this relationship.

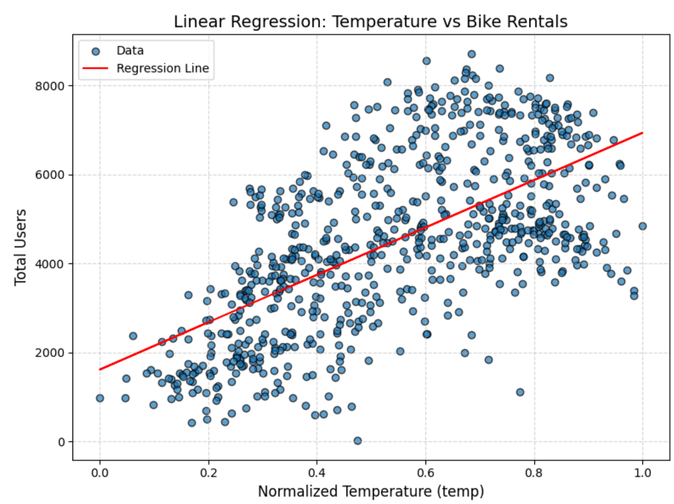

Figure 6 Linear Regression: Temperature vs Bike Rentals

Key metrics:

- **R^2 Score**: Indicates a strong linear relationship between temperature and rentals.
- **Mean Squared Error (MSE)**: Measures the error of predictions.

Future Prediction

The regression model was extended to predict bike rentals for hypothetical future temperature values.

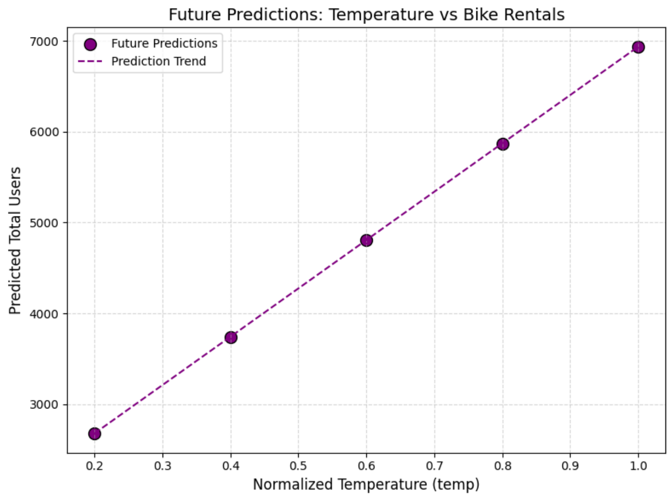

Figure 7 Future Predictions: Temperature vs Bike Rentals

Key insights:

- Predictions show a linear increase in bike rentals as temperature rises, validating the model's reliability for future estimates.

Conclusion

This report highlights key findings:

1. Seasonality and weather strongly influence bike rental patterns.
2. Temperature positively impacts rental counts, as evidenced by regression analysis.
3. Clustering provided meaningful segmentation for understanding user behavior.
4. Predictive analytics can aid in forecasting future rental demands, offering actionable insights for resource planning.

GitHub Repository

The full codebase, along with reproducible Python scripts and notebooks, is available at: Click Here