## Capital Bikeshare data analysis project

How to properly predict bike rental demand

**Jorge Villatoro**

930906132

**12/10/2024**

# Project Section 1: Introduction

Dataset: [Bike Sharing](http://archive.ics.uci.edu/dataset/275/bike+sharing+dataset)

Accomplishment Schedule

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **Week** | **Dates** | **Objectives** | **Deliverables** | | Week 1 | 09/30 - 10/06 | Set up project lab environment; select dataset | Confirm dataset; initial review | | Week 2 | 10/07 - 10/13 | Draft Sections 1 and 2 | Drafts of Section 1: Introduction, Section 2: Context | | Week 3 | 10/14 - 10/20 | Define the problem | Draft of Section 3: Defining the Problem | | Week 4 | 10/21 - 10/27 | Data collection and preparation | Progress on Section 4: Data | | Week 5 | 10/28 - 11/03 | Data analysis | Formative submission (Sections 1-5) | | Week 6 | 11/04 - 11/10 | Model evaluation | Complete Section 6: Evaluation | | Week 7 | 11/11 - 11/17 | Interpret results | Draft of Section 7: Interpretation | | Week 8 | 11/18 - 11/24 | Communicate findings | Sections 8: Recommendations, 9: Executive Summary | | Week 9 | 11/25 - 12/01 | Refinement and presentation prep | Finalize all project components; prepare for presentation | | Week 10 | 12/02 - 12/08 | Finalization and submission | Finalizing presentation, and presentation prep | | Week 11 | 12/09 - 12/15 | Presentation of findings | Submit complete project | |

# Project Section 2: CONTEXT

The Bike Sharing dataset focuses on the Capital Bikeshare system in the Washington, D.C., area, covering bike rentals from 2011 to 2012. Developed by the Laboratory of Artificial Intelligence and Decision Support at the University of Porto under the guidance of Hadi Fanaee-T, the dataset was compiled from public sources and supplemented with weather data from external sources. This data includes factors such as temperature, humidity, wind speed, seasonal information, holidays, and whether the day is a working day. According to the dataset description, the data supports two major tasks: regression analysis and anomaly detection.

In this analysis, I plan to conduct Exploratory Data Analysis (EDA) to understand general trends, seasonality, and the distribution of bike rentals over time. Additionally, I will be modeling the data, leveraging regression analysis techniques I learned in previous projects using Excel. This approach will allow me to predict rental behaviors based on weather and calendar variables, identifying key predictors. By applying these methods, I aim to uncover insights into factors influencing bike rentals in Washington, D.C., during 2011-2012.

List of critical stakeholders to Capital Bikeshare:

|  |  |
| --- | --- |
| **Stakeholder** | **Why They’re Important** |
| **Bike Users (Customers)** | The end-users of the service who rely on the availability of bikes for commuting, leisure, and errands. Their satisfaction determines the success of the bikeshare system. |
| **City Planners** | Utilize bike rental data to plan infrastructure improvements, such as bike lanes and docking stations, enhancing urban mobility and sustainability. |
| **Operations Team** | Responsible for maintaining and redistributing bikes to ensure stations are neither overstocked nor empty, ensuring smooth operations. |
| **Local Businesses** | Depend on bike traffic to attract customers, especially in areas with high foot and bike traffic. Increased bike use can support economic activity. |
| **Government Agencies** | Oversee funding, regulatory compliance, and policymaking, particularly for sustainability and urban mobility initiatives tied to bikeshare programs. |
| **Environmental Groups** | Advocate for the bikeshare system as a tool to reduce carbon emissions and promote green transportation alternatives. |
| **Data Scientists** | Analyse rental patterns, weather impacts, and operational inefficiencies to optimize the system and forecast demand effectively. |
| **Marketing Team** | Use customer data and trends to design campaigns that attract new users, retain existing ones, and promote special offers or events. |
| **Technology Providers** | Provide the software and systems that manage bike rentals, payments, and station operations, ensuring a seamless user experience. |
| **Community Groups** | Advocate for equitable access to bikes across all neighbourhoods, ensuring underserved areas are included in bike distribution plans. |

**7-Step Approach for Solving Capital Bikeshare’s Problem**

1. **Define the Problem**
   * **Identify the Issue:** Understand the core challenge of ensuring optimal bike availability to meet demand across seasons and weather conditions.
   * **Articulate the Problem:** Clearly state the issue of balancing supply to avoid understocked or overstocked bike stations.
   * **Set Objectives:** Define success as achieving operational efficiency, reducing wasted resources, and enhancing customer satisfaction.
2. **Gather the Right Data**
   * **Identify Data Sources:** Utilize bike rental data, weather records, and seasonal information.
   * **Assess Data Quality:** Ensure the dataset is complete, accurate, and reliable for analysis.
   * **Collect and Organize Data:** Compile data in a structured format while ensuring compliance with privacy standards.
3. **Analyze the Data**
   * **Select Analytical Techniques:** Choose regression models, time-series methods, and visualization tools to explore trends.
   * **Process Data:** Clean and transform the dataset, addressing outliers and missing values.
   * **Conduct Analysis:** Uncover patterns in bike rentals based on temperature, humidity, seasonality, and other variables.
4. **Evaluate Findings**
   * **Assess Validity:** Verify that the analytical methods are sound and produce credible results.
   * **Check Reliability:** Ensure findings are consistent and replicable with similar datasets.
   * **Interpret Relevance:** Relate the results back to the operational problem of optimizing bike supply.
5. **Interpret Results**
   * **Understand Implications:** Analyze how weather and seasonality influence bike rentals and identify actionable insights.
   * **Draw Conclusions:** Summarize the role of key predictors, such as temperature and humidity, in determining demand.
   * **Acknowledge Limitations:** Highlight potential confounding factors and areas requiring further analysis.
6. **Communicate Insights**
   * **Prepare Documentation:** Develop reports, presentations, and visual aids that summarize key findings.
   * **Tailor Communication:** Adjust the message to meet the needs of executives, operations teams, and other stakeholders.
   * **Engage Stakeholders:** Present findings, encourage feedback, and discuss implementation strategies.
7. **Decide and Act**
   * **Formulate Options:** Propose solutions, such as dynamic bike redistribution based on weather forecasts.
   * **Evaluate Feasibility:** Weigh the costs, risks, and benefits of each option.
   * **Implement Recommendations:** Choose the most practical and impactful course of action to optimize operations and meet business objectives.

# Project Section 3: Defining the problem

**1. Problem Definition:**

To predict how many bikes people will need, and when.

**2. Question Definition and link with the problem:**

How many bikes are needed at each station to satisfy daily demand across different seasons and weather conditions? The company needs to know how many bikes are enough at each bike station to ensure that everyone has a bike when they need it. The problem here is that bike rentals are seasonal. Having a look at the data, it I clear that there are more bike rentals in the summer months, likely because it Is warm and sunny out. So, my job as an analyst is to identify the optimal number of bikes needed at each station during slow and peak rental months/days.

**3. Explanation of the business relevance of the problem**

Meeting demand without oversupply is crucial for Capital Bikeshare’s operational efficiency and customer satisfaction. If stations regularly run empty or have excess bikes, it leads to customer frustration, lost revenue, and wasted resources in redistribution efforts. We need to find a suitable number of supply to meet demand. I plan on doing so by analyzing historical data and predicting demand based on weather, season, and time of day. I can use the data to forecast rental peaks, helping with bike rebalancing so that we can minimize the amount of empty or overfull docking stations. This will also allow for optimized staffing and maintenance planning, reducing operating costs and ensuring the company’s resources are focused on periods and locations with the highest demand.

# Project Section 4: Data

**Dataset Description**

The datasets **day.csv** and **hour.csv** provide a detailed look at bike-sharing activity over different time intervals, which is essential for understanding seasonal and weather impacts on demand. Here’s a breakdown of the datasets:

1. **Day.csv**: This dataset contains 731 daily records, with variables such as:
   * **Date** (dteday) and other time indicators (season, yr, mnth, weekday, holiday, workingday).
   * **Weather conditions** (weathersit, temp, atemp, hum, windspeed).
   * **Usage metrics** (casual, registered, cnt), where cnt is the total number of rentals.
2. **Hour.csv**: With 17,379 hourly records, this dataset includes the same variables as **day.csv** but adds an hr column to capture hourly data. The structure and types are similar, with a mixture of integers, floats, and date strings.

**Exploratory Data Analysis (EDA)**

To address the project’s goal — accurately forecasting bike demand by season and weather — I focused on the following variables:

* **cnt**: The total count of daily rentals, which directly indicates demand. The mean daily rentals hover around 4504, with a wide range from 22 to 8714, indicating strong seasonal and weather-related variability.
* **temp** and **atemp**: Both actual and perceived temperatures have mean values around 0.5, reflecting normalized scales. Warmer temperatures correlate with higher rentals, as expected.
* **hum** (humidity) and **windspeed**: These environmental factors affect comfort, which is likely to influence rental patterns.

A plot of cnt over time highlights visible seasonal peaks, particularly during warmer months, suggesting temperature and other seasonal variables as key predictors.

These attributes were selected based on their significant impact on predicting bike rental demand and optimizing operations, aligning with the objectives of Capital Bikeshare.

**Summary Statistics Table:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Std Dev** | **Min** | **Max** |
| cnt | 189.46 | 181.39 | 1 | 977 |
| temp | 0.5 | 0.19 | 0.02 | 1 |
| hum | 0.63 | 0.19 | 0 | 1 |
| windspeed | 0.19 | 0.12 | 0 | 0.85 |
| hr | 11.55 | 6.91 | 0 | 23 |

A screenshot of a graph

Description automatically generated

**Is the Data Right for the Problem?**

I believe the data provided works well for regression analysis, or time series analysis to predict demand. There are a few things I would consider:

* I think considering outliers for variables such as weather, or holidays is important to look at. These outliers could skew data if we don’t handle them. No outliers will ensure a robust model.
* I also analyzed the other variables alongside temperature such as humidity, and windspeed. These shows little to no relation to the number of bikes rented. Seems as though weather has a greater affect on bike rentals, giving me the idea that bike rentals are very much seasonal.

**Additional Plots:**

**Average Bike Rentals by Hour of the Day:**

**A graph of a bike rental

Description automatically generated**

* Shows clear peaks during typical commuting hours (around 8 AM and 6 PM).
* Indicates the role of daily commuting patterns in bike rental usage.

** Bike Rentals by Season:A graph showing different colored squares

Description automatically generated**

* Highlights higher rentals during summer and fall, likely due to more favorable weather conditions.
* Winter having more rentals than Spring is surprising, and important to keep note of…

**Average Rentals Depending on Temperature**



* Temperature has a clear correlation to the number of bikes are being rented out.
* This chart indicates key factors such as seasonality and of course temperature. `

# Project Section 5: Analysis

In this analysis, the outcome variable we’re focusing on is **bike rental count (cnt)**. The primary question we aim to answer is: *How do weather conditions affect bike rental demand over time?* This inquiry is directly tied to an operational problem for Capital Bikeshare, where understanding the patterns in bike rentals based on weather can help optimize bike distribution, enhance customer satisfaction, and reduce costs related to underutilization or over-supply of bikes. By predicting bike rental demand, we can ensure that each station maintains an adequate supply relative to anticipated demand, particularly during peak seasons.

**Outcome Variable and Predictor Variables**

The outcome variable is **bike rental count (cnt)**, measured as the daily total bike rentals across all stations in the system. To predict this outcome, we’ve selected **two main predictor variables** based on their hypothesized impact on bike rental behavior: **temperature (temp)** and **humidity (hum)**.

Temperature, measured as a scaled variable, is expected to have a positive correlation with bike rentals, particularly in moderate-to-high temperature ranges where biking is comfortable. Conversely, humidity is anticipated to have an inverse relationship, where higher humidity may discourage bike rentals due to discomfort or potential precipitation.

**Models for Prediction**

We developed two models to predict bike rentals (cnt) using different combinations of variables, both evaluated over the entire dataset.

1. **Model 1: Linear Regression with Temperature and Count**
   * This model explores the relationship between bike rental count and temperature. We hypothesize that as temperature increases, bike rentals will also increase, especially in favorable conditions.
   * **Formula**: cnt=β0+β1⋅temp+ϵcnt=β0​+β1​⋅temp+ϵ
   * Here, cnt is the dependent variable (bike rentals), temp is the independent variable (temperature), and β0β0​and β1β1​ are the intercept and slope coefficients, respectively. ϵ represents the error term.
2. **Model 2: Multiple Linear Regression with Temperature, Humidity, and Count**
   * This model expands on Model 1 by incorporating humidity as an additional predictor. By adding humidity, we can observe the combined effects of both temperature and comfort level on bike rental demand.
   * **Formula: cnt**=β0+β1⋅temp+β2⋅hum+ϵcnt=β0​+β1​⋅temp+β2​⋅hum+ϵ
   * In this model, both temp and hum are independent variables, allowing us to analyze their joint impact on the bike rental count.

**Analysis in Excel**

To conduct these analyses in Microsoft Excel, I used the **Data Analysis Toolpak** for regression analysis, which allows for efficient computation of regression coefficients and statistical measures such as R-squared and p-values.

**Model 1 Analysis: Temperature as a Predictor of Bike Rentals**

After running a simple linear regression in Excel with cnt as the dependent variable and temp as the independent variable, we observed a significant positive correlation between temperature and bike rentals. The resulting R-squared value of 0.62 indicates that approximately 62% of the variance in bike rentals is explained by temperature alone. This suggests that temperature is a meaningful predictor of bike rental demand, although other factors also contribute to variations.

* **Table 1**: Summary of Regression Output for Model 1
* A screenshot of a spreadsheet

  Description automatically generated

**Plot 1**: Scatter Plot with Trendline for Temperature vs. Bike Rentals

* + This plot shows the linear relationship between temperature and bike rental count, with the trendline indicating an upward slope.

**The regression equation is:**

Predicted Rentals (cnt) = 2657.89512 - 2492.8541 × Hour Average + 6886.97373 × Temperature Average

**What Does This Mean?**

**Intercept (2657.89512):**

* + This is the baseline number of rentals when the hour and temperature averages are both zero. While this isn’t a real-world scenario (since both averages wouldn't be zero), it serves as a starting point in the model.

**Hour Average (-2492.8541):**

* + For every 1-unit increase in the hourly average, the predicted rentals decrease by about 2493.
  + This means bike rentals are lower during certain hours, especially late at night or early morning when fewer people are using bikes.

**Temperature Average (+6886.97373):**

* + For every 1-unit increase in the temperature average (which is normalized between 0 and 1), bike rentals increase by about 6887.
  + This suggests a strong relationship between temperature and bike rentals—warmer days lead to significantly higher demand for bikes.

**Model 2 Analysis: Temperature and Humidity as Predictors of Bike Rentals**

For Model 2, I ran a multiple linear regression using cnt as the dependent variable and both temp and hum as independent variables. The R-squared value for this model increased to 0.65, indicating that 65% of the variance in bike rentals is explained by the combination of temperature and humidity. The inclusion of humidity improved the model’s explanatory power, suggesting that humidity levels, along with temperature, significantly influence rental demand.

* **Table 2**: Summary of Regression Output for Model 2

A screenshot of a spreadsheet

Description automatically generated

* **Plot 2**: Multiple Regression Line Plot for Temperature, Humidity, and Bike Rentals
  + This plot illustrates the predicted bike rental count based on varying levels of temperature and humidity. Higher temperatures coupled with lower humidity levels generally correspond to higher bike rentals, reinforcing the idea that favorable weather conditions encourage biking.

Predicted Rentals (cnt) = 2657.89512 - 2492.8541 × Hour Average + 6886.97373 × Temperature Average

**What Does This Mean?**

**Intercept (2657.89512):**

* + This is the baseline number of bike rentals when all other factors (hour and temperature) are at their averages. It serves as the starting point for predictions.

**Temperature Average (+6886.97373):**

* + For every 1-unit increase in the normalized temperature (which ranges from 0 to 1), bike rentals increase by about 6887. This highlights the significant positive impact of temperature on bike rentals. Warmer days encourage more people to rent bikes.

**Humidity Average (-2492.8541):**

* + For every 1-unit increase in normalized humidity, bike rentals decrease by about 2493. Higher humidity levels can make outdoor activities less comfortable, leading to fewer rentals.



**Model 1 Vs. Model 2**

From the analyses, both temperature and humidity are shown to significantly affect bike rental demand, with temperature being a strong positive predictor and humidity having a negative impact. Model 2, which incorporates both variables, is more robust and explains a greater portion of the variance in bike rentals. These insights could inform operational decisions, such as adjusting bike supply based on anticipated weather conditions, ultimately improving the efficiency and reliability of Capital Bikeshare’s services.

Model 3: Bike Rentals Considering Windspeed

A spreadsheet with numbers and a number

Description automatically generated

**Predicted Rentals (cnt) = 163.185332 + 138.232962 × Windspeed**

**What Does This Mean?**

**Intercept (163.185332):**

* + This is the baseline number of bike rentals when windspeed is zero. It represents the starting point for predictions when there is no wind.

**Windspeed (138.232962):**

* + For every 1-unit increase in normalized windspeed (from 0 to 1), bike rentals increase by about **138.23**. This suggests a slight positive correlation between windspeed and bike rentals.

**Insights:**

* **R-Squared (0.0087):**
  + This means that only **0.87%** of the variability in bike rentals is explained by windspeed. This is a very low number, indicating that windspeed is not a strong predictor of bike rentals.
* **Statistical Significance:**
  + Despite the weak predictive power, the p-value is extremely small (7.34E-35), suggesting that windspeed does have a statistically significant, yet minimal, relationship with bike rentals.
* **The Plot:**
  + The scatter plot with a line fit shows that rentals are distributed quite evenly across different windspeed levels, with no clear pattern. This reinforces the idea that windspeed alone doesn’t strongly influence bike rental behavior.

I tried using windspeed to predict bike rentals, and while it technically has a statistically significant effect, it’s not a strong one. The R-squared value is very low (less than 1%), which means windspeed doesn’t explain much of the variation in bike rentals. The scatter plot backs this up because the data points are all over the place, and the fitted line is nearly flat.

In short, windspeed might play a small role in how many bikes are rented, but it’s almost insignificant compared to temperature and even humidity.

# Project Section 6: Evaluation

**Confounding: Alternative Explanations for Observed Relationships**

First, it’s likely that seasonality plays a big role here. Warmer months often bring holidays, outdoor events, and vacation periods. These naturally increase overall activity levels, and bike rentals are probably just one part of that trend. So, while it looks like warmer weather is driving demand, it might be these other seasonal factors causing the spike.

Second, urban activity during warmer weather is another potential confounder. Think about how tourists and locals flock to outdoor events or attractions when the weather is nice. This boost in activity might increase bike rentals, but not because of the temperature itself. It’s more about the opportunities that come with warmer conditions.

These factors—seasonality and urban activity—aren’t directly accounted for in my models. That means the models might be overstating the impact of temperature and humidity. It’s important to keep these confounders in mind when interpreting the results. By doing so, I can provide more balanced insights and avoid drawing conclusions that oversimplify the relationships in the data.

**Overfitting: Data Sufficiency and Model Complexity**

The risk of overfitting is minimal given the size and granularity of the datasets used. The daily dataset contains 731 records, while the hourly dataset includes over 17,000 observations, providing a strong foundation for the analysis. The models themselves are relatively simple, with Model 1 focusing solely on temperature and Model 2 adding humidity as an additional variable. However, these straightforward models may not fully capture more complex patterns, such as hourly variations in demand during peak commuting times. To address this, future analyses could incorporate time series methods to account for the hourly variations. Doing so could help show a more detailed picture in the hourly timeframe. An average hourly Vs. Day plot was added to analyze the overall average rentals in a day’s time; but no hourly data was inserted into any models. Also, Cross-validation could also confirm the stability of the models across different subsets of the data.

**Causality: Correlation vs. Causation**

The models demonstrate a strong correlation between weather variables and bike rentals, but they do not confirm any kind of causation. While temperature and humidity are important predictors, they do not directly cause increased bike rentals. Other unobserved factors, such as holidays, grand openings of new businesses or parks, or economic conditions of the country, could also influence rental activity. Therefore, the findings should be interpreted as correlations that provide useful insights for forecasting rather than solidified causal relationships. For example, the relationship between temperature and rentals could be driven by increased outdoor activities during warmer weather, but this does not mean temperature alone dictates biking preferences. Future research can be done to explore causality through experimental designs or by incorporating additional control variables, such as marketing data or urban activity metrics.

**Significance and Effect Size: Business Implications**

The bike data findings carry significant possible business advantages for Capital Bikeshare. By leveraging the correlation between weather conditions and rental demand, the company can restock bike docking stations, reducing the likelihood of empty or overcrowded stations. This operational advantage not only minimizes resource wastage but also enhances customer satisfaction, potentially increasing user retention and revenue for the business. The effect sizes of the predictors further supports these insights. Temperature, as a single variable, has a substantial effect size, as evidenced by the high R-squared value (0.62) in Model 1. This underscores its practical and reliable use in forecasting demand. Humidity, on the other hand, provides a smaller effect size, contributing only a marginal improvement in Model 2. While the addition of the humidity variable is helpful, its impact is negligible compared to the impact of the temperature variable.

**Further Comment**

The models developed for analysing bike rental demand are highly effective, but Model 1, which uses temperature as the sole predictor, is the most practical choice moving forward. With an R-squared of 0.62, it captures a substantial portion of the variance in bike rental demand while maintaining simplicity and ease of interpretation. Although Model 2 offers a slight improvement with an R-squared of 0.65 by adding humidity, the incremental gain is marginal and doesn’t justify the added complexity. By focusing solely on temperature, Model 1 remains straightforward, making it ideal for operational decision-making and resource allocation without introducing unnecessary variables. Introducing additional predictors, such as humidity or windspeed may lead to diminishing returns in accuracy and could complicate implementation for stakeholders who prioritize cost-effective actionable insights over intricate models. Additionally, the dataset used ensures the reliability of this simpler model while minimizing overfitting risks. Given its balance of clarity, efficiency, and significant business implications, Model 1 is the optimal choice for Capital Bikeshare's forecasting and operational needs.

# Project Section 7: Interpretation

**Restating the Original Question**

The original question guiding this analysis was: "How many bikes are needed at each station to satisfy daily demand across different seasons and weather conditions?" This question is the central idea optimizing Capital Bikeshare's operations, balancing the availability of bikes with customer demand, and minimizing instances of overstocked or understocked stations. Addressing this issue helps the company improve operational efficiency, enhance customer satisfaction, and reduce resource wastage. Both models aimed to forecast bike rental demand by examining weather conditions, with the goal of providing operational business insights to align bike distribution with observed demand data patterns.

**Assessing Directionality, Magnitude, and Uncertainty**

**Model 1 (Temperature as Predictor):**  
Model 1 shows a strong positive relationship between temperature and bike rentals. As temperatures rise, especially into moderate-to-warm ranges, demand for bikes increases significantly. The model explains about 62% of the variance in bike rentals, which makes temperature a reliable and important factor for predicting demand. However, it doesn’t account for other possible influences like the time of day or special events, which could create some uncertainty. Even with these limitations, the simplicity of Model 1 makes it easy to understand and apply, making it a very actionable tool for forecasting.

**Model 2 (Temperature and Humidity as Predictors):**  
Model 2 builds on Model 1 by adding humidity as a factor. It shows that higher humidity slightly reduces bike rentals, likely because it makes outdoor activities less comfortable. While temperature remains the strongest predictor, adding humidity improves the model’s accuracy a bit, increasing the R-squared value to 65%. Still, this small gain doesn’t outweigh the fact that humidity has a weaker influence compared to temperature. Like Model 1, this model doesn’t include variables like windspeed or whether it’s a weekday or weekend, which leaves some room for improvement. Overall, while Model 2 provides a little more detail, its added complexity doesn’t make it significantly more useful than Model 1.

**Overall Interpretation of the Final Model**

Model 1, which focuses only on temperature, ended up being the best choice because it’s simple and practical. It clearly showed important patterns in bike rental demand, with temperature standing out as the strongest predictor. Bike rentals tend to peak on warmer days and drop off during colder seasons, which makes a lot of sense. The seasonal trends were also clear rentals were highest during summer and lowest in spring. While Model 2 added humidity as a secondary factor and slightly improved accuracy, the added complexity wasn’t worth it for practical use. Model 1 keeps things straightforward and still delivers the key insights needed for effective decision-making.

**Implications of the Model**

Model 1 connects well with Capital Bikeshare’s main challenge and provides practical insights the company can act on. It’s a dependable way to forecast bike rental demand across different seasons, which helps the company make smarter decisions about where and when to move bikes. This means stations are less likely to be overstocked or run out of bikes, cutting costs and making better use of resources. On top of that, being able to predict demand improves customer satisfaction—people are less likely to find empty stations during busy times or too many bikes during slower periods. These predictions also help with bigger-picture planning, like adjusting marketing, resource allocation, and even infrastructure investments to align with seasonal trends. Overall, Model 1 is simple, reliable, and directly relevant to solving Capital Bikeshare’s operational challenges, making it a great tool for the job.

# Project Section 8: Recommendations

After analysing the Capital Bikeshare dataset, I have identified several actionable recommendations that can greatly improve bike rental operations, enhance customer satisfaction, and optimize overall efficiency.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Approach | Details | Data to Back Up Recommendation | Benefits | Risks |
| 1. Implement Temperature-Based Forecasting | Use a temperature-focused model to predict bike rental demand and adjust bike distribution accordingly. | Model 1 explains 62% of bike rental variability using temperature alone. | Improves operational efficiency by ensuring stations are adequately stocked based on demand. | Over-reliance on temperature could miss other important factors like time of day or special events. |
| 2. Develop Seasonal Redistribution Plans | Increase bike supply during warmer months and reduce it during colder months to match seasonal demand patterns. | Seasonal analysis shows higher rentals in summer and significantly lower rentals in winter. | Reduces excess bikes during low-demand periods, lowering operational costs and resource waste. | Risk of undersupplying bikes during unseasonal weather patterns or unexpected demand spikes. |
| 3. Integrate Humidity as a Secondary Factor | Add humidity to the model for a more nuanced prediction of bike rental demand. | Model 2 increases accuracy (R² = 0.65) by incorporating humidity, revealing a slight negative impact on demand. | Provides a deeper understanding of environmental impacts on rentals, enhancing forecasting precision. | Added complexity may not justify the marginal improvement, potentially complicating operations. |
| 4. Focus on Hourly Demand Patterns | Use hourly data to refine bike allocation during peak commuting times and quiet hours. | Hourly data reveals distinct peaks during morning and evening commute times (8 AM and 6 PM). | Ensures better bike availability during peak hours, improving customer satisfaction. | Requires more frequent redistribution, which could increase labor and logistics costs. |

**Recommendation for Capital Bikeshare Executives**

After analyzing the 2011-2012 dataset, I recommend Capital Bikeshare move forward using Model 1, which highlights temperature as the sole predictor for bike rental demand. This model balances simplicity with reliability, providing actionable insights without overcomplicating the analysis. Based on Model 1, the company should adjust bike distribution dynamically by season. Specifically, increase bike availability during warmer months, especially on weekends and holidays, when demand peaks. Conversely, reduce bike supply during colder months to lower operational costs and reallocate resources toward maintenance and system upgrades. Additionally, implementing a temperature-based forecasting system can further enhance efficiency by predicting rental demand for warmer days and triggering proactive bike redistribution to high-demand stations. By focusing on temperature, Capital Bikeshare can confidently address the primary driver of demand while maintaining operational clarity.

**Ethical, Privacy, and Governance Considerations**

To ensure these recommendations align with ethical and governance standards, Capital Bikeshare should prioritize data privacy by anonymizing user data and avoiding the collection of unnecessary personal information. Clear communication with customers about how their data supports operational improvements is essential for maintaining trust and goodwill. Furthermore, the company should use these insights to promote equitable bike distribution, ensuring all neighborhoods, including underserved areas, have fair access to the service. This prevents resource concentration in higher-demand locations at the expense of equity. By adopting these practices, Capital Bikeshare can improve customer satisfaction and operational efficiency while fostering trust and demonstrating a commitment to ethical growth.

# Project Section 9: EXECUTIVE SUMMARY

**Executive Summary**

The Capital Bikeshare system plays a vital role in Washington, D.C.’s urban transportation network, providing a sustainable and convenient option for residents and visitors. To further enhance its operational efficiency, I analyzed the 2011-2012 Capital Bikeshare dataset to identify the key factors influencing bike rental demand and provide actionable insights for improvement. The goal of this project was to understand the patterns behind bike rentals, focusing on weather conditions and seasonal trends, and to use these findings to recommend strategies for optimizing bike availability across the city.

This analysis relied on a dual-model approach, with Model 1 focusing solely on temperature and Model 2 incorporating both temperature and humidity. Model 1 demonstrated a strong positive relationship between temperature and bike rentals, showing that warmer days significantly increase demand, particularly in the summer. This model explained 62% of the variance in bike rentals, providing a clear and actionable foundation for forecasting. Model 2 added humidity as a secondary variable, revealing that higher humidity slightly reduces demand. While this addition improved the accuracy to 65%, the incremental gain did not justify the added complexity for practical use.

**A graph of a bike rental

Description automatically generated**

Key findings included the importance of temperature as the primary predictor of bike rentals, with demand peaking during warm summer days and dropping during colder months. Seasonal trends were also clear, with summer seeing the highest rentals and winter experiencing the lowest. Hourly data further highlighted distinct peaks during commuting hours, around 8 AM and 6 PM, demonstrating the influence of daily routines on rental patterns.

Based on these findings, I recommend Capital Bikeshare implement a temperature-based forecasting system to optimize bike redistribution efforts, ensuring stations are adequately stocked during high-demand periods and scaling back during low-demand seasons. Additionally, creating seasonal redistribution plans can further improve operational efficiency by aligning resources with seasonal trends. While humidity offers additional insights, it can be considered a secondary factor to refine predictions where necessary.

This project not only highlights the key drivers of bike rental demand but also provides practical strategies to improve the system’s efficiency, reduce costs, and enhance customer satisfaction. By leveraging these data-driven insights, Capital Bikeshare can better align its operations with user needs and support sustainable transportation goals in Washington, D.C.

# Project Section 10: Presentation

[Presentation link](https://youtu.be/s-9HfkgCLFI?si=GK_eD1hv9Oj4viRI)

A person riding a bicycle

Description automatically generatedA diagram of a bicycle

Description automatically generated

A white and black card with black text

Description automatically generatedA person riding a bicycle

Description automatically generated