

# Forecasting Bikeshare Demand

Using Time Series Models in R



Seif H. Kungulio

# Forecasting Bikeshare Demand Using Time Series Analysis

Seif H. Kungulio

December 19, 2025



## Contents

<b>Business Understanding</b>	<b>1</b>
Introduction . . . . .	1
Problem Statement . . . . .	1
Objectives . . . . .	1
Key Stakeholders . . . . .	1
Project Scope . . . . .	1
Success Criteria . . . . .	1
<b>Data Understanding</b>	<b>2</b>
Dataset Overview . . . . .	2
Data Loading . . . . .	2
Data Exploration . . . . .	4
Initial Findings . . . . .	4
Summary of Insights . . . . .	4
<b>Data Preparation</b>	<b>5</b>
Data Cleaning . . . . .	5
Feature Engineering . . . . .	5
Data Transformation . . . . .	6
Final Dataset Summary . . . . .	6
Data Quality Report . . . . .	6
Data Dictionary . . . . .	6
Summary of Preparation Steps . . . . .	6
<b>Modeling</b>	<b>7</b>
Modeling Approach . . . . .	7
Model Development . . . . .	7
Model Evaluation . . . . .	7
Model Selection . . . . .	7
Summary of Modeling Steps . . . . .	7
<b>Evaluation</b>	<b>8</b>
Results Summary . . . . .	8
Business Impact Assessment . . . . .	8
Lessons Learned . . . . .	8
Recommendations . . . . .	8

<b>Deployment</b>	<b>9</b>
Deployment Plan . . . . .	9
Monitoring and Maintenance Strategy . . . . .	9
Future Work Suggestions . . . . .	9
Summary of Deployment Steps . . . . .	9
Acknowledgments . . . . .	9
About the Author . . . . .	9
<b>Conclusion</b>	<b>10</b>
Project Summary . . . . .	10
Key Takeaways . . . . .	10
Final Thoughts . . . . .	10
<b>References</b>	<b>12</b>
List of References . . . . .	12
Citation Style Used . . . . .	12
Additional Reading Materials . . . . .	12
Acknowledgment of Sources . . . . .	12
<b>Appendix</b>	<b>13</b>

## Business Understanding

Introduction

Problem Statement

Objectives

Key Stakeholders

Project Scope

Success Criteria

## Data Understanding

### Dataset Overview

#### Data Loading

```
day_raw <- read_csv("resources/data/day.csv")
```

```
## Rows: 731 Columns: 16
## -- Column specification -----
## Delimiter: ","
## dbl (15): instant, season, yr, mnth, holiday, weekday, workingday, weathers...
## date (1): dteday
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(day_raw)
```

```
## Rows: 731
## Columns: 16
## $ instant      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, ~
## $ dteday       <date> 2011-01-01, 2011-01-02, 2011-01-03, 2011-01-04, 2011-01-05~
## $ season        <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ yr           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ mnth          <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ holiday        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ weekday        <dbl> 6, 0, 1, 2, 3, 4, 5, 6, 0, 1, 2, 3, 4, 5, 6, 0, 1, 2, 3, 4, ~
## $ workingday     <dbl> 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, ~
## $ weathersit     <dbl> 2, 2, 1, 1, 1, 2, 2, 1, 1, 1, 2, 1, 2, 2, 2, 2, ~
## $ temp           <dbl> 0.3441670, 0.3634780, 0.1963640, 0.2000000, 0.2269570, 0.20~
## $ atemp          <dbl> 0.3636250, 0.3537390, 0.1894050, 0.2121220, 0.2292700, 0.23~
## $ hum            <dbl> 0.805833, 0.696087, 0.437273, 0.590435, 0.436957, 0.518261, ~
## $ windspeed       <dbl> 0.1604460, 0.2485390, 0.2483090, 0.1602960, 0.1869000, 0.08~
## $ casual          <dbl> 331, 131, 120, 108, 82, 88, 148, 68, 54, 41, 43, 25, 38, 54~
## $ registered      <dbl> 654, 670, 1229, 1454, 1518, 1518, 1362, 891, 768, 1280, 122~
## $ cnt             <dbl> 985, 801, 1349, 1562, 1600, 1606, 1510, 959, 822, 1321, 126~
```

```
hour_raw <- read_csv("resources/data/hour.csv")
```

```
## Rows: 17379 Columns: 17
## -- Column specification -----
## Delimiter: ","
## dbl (16): instant, season, yr, mnth, hr, holiday, weekday, workingday, weat...
## date (1): dteday
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(hour_raw)
```

```
## Rows: 17,379
## Columns: 17
## $ instant    <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, ~
## $ dteday     <date> 2011-01-01, 2011-01-01, 2011-01-01, 2011-01-01~
## $ season     <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ yr         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ mnth        <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ hr          <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1~
## $ holiday     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ weekday     <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, ~
## $ workingday  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ weathersit   <dbl> 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, ~
## $ temp         <dbl> 0.24, 0.22, 0.22, 0.24, 0.24, 0.24, 0.22, 0.20, 0.24, 0.32, ~
## $ atemp        <dbl> 0.2879, 0.2727, 0.2727, 0.2879, 0.2879, 0.2576, 0.2727, 0.2~
## $ hum          <dbl> 0.81, 0.80, 0.80, 0.75, 0.75, 0.75, 0.80, 0.86, 0.75, 0.76, ~
## $ windspeed    <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0896, 0.0000, 0.0~
## $ casual       <dbl> 3, 8, 5, 3, 0, 0, 2, 1, 1, 8, 12, 26, 29, 47, 35, 40, 41, 1~
## $ registered   <dbl> 13, 32, 27, 10, 1, 1, 0, 2, 7, 6, 24, 30, 55, 47, 71, 70, 5~
## $ cnt          <dbl> 16, 40, 32, 13, 1, 1, 2, 3, 8, 14, 36, 56, 84, 94, 106, 110~
```

```
skimr::skim(day_raw)
```

Table 1: Data summary

Name	day_raw
Number of rows	731
Number of columns	16
Column type frequency:	
Date	1
numeric	15
Group variables	
	None

### Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
dteday	0	1	2011-01-01	2012-12-31	2012-01-01	731

### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
instant	0	1	366.00	211.17	1.00	183.50	366.00	548.50	731.00	
season	0	1	2.50	1.11	1.00	2.00	3.00	3.00	4.00	
yr	0	1	0.50	0.50	0.00	0.00	1.00	1.00	1.00	
mnth	0	1	6.52	3.45	1.00	4.00	7.00	10.00	12.00	

skim__variables	len_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
holiday	0	1	0.03	0.17	0.00	0.00	0.00	0.00	1.00	
weekday	0	1	3.00	2.00	0.00	1.00	3.00	5.00	6.00	
workingday	0	1	0.68	0.47	0.00	0.00	1.00	1.00	1.00	
weathersit	0	1	1.40	0.54	1.00	1.00	1.00	2.00	3.00	
temp	0	1	0.50	0.18	0.06	0.34	0.50	0.66	0.86	
atemp	0	1	0.47	0.16	0.08	0.34	0.49	0.61	0.84	
hum	0	1	0.63	0.14	0.00	0.52	0.63	0.73	0.97	
windspeed	0	1	0.19	0.08	0.02	0.13	0.18	0.23	0.51	
casual	0	1	848.18	686.62	2.00	315.50	713.00	1096.00	3410.00	
registered	0	1	3656.17	1560.26	20.00	2497.00	3662.00	4776.50	6946.00	
cnt	0	1	4504.35	1937.21	22.00	3152.00	4548.00	5956.00	8714.00	

## Data Exploration

### Initial Findings

### Summary of Insights

## Data Preparation

## Data Cleaning

```
day_raw %>%
  summarise(across(everything(), ~ sum(is.na(.)))) %>%
  pivot_longer(everything(), names_to = "variable", values_to = "missing_count") %>%
  arrange(desc(missing_count)) %>%
  kable(caption = "Missing Values in Day Dataset")
```

Table 4: Missing Values in Day Dataset

variable	missing_count
instant	0
dteday	0
season	0
yr	0
mnth	0
holiday	0
weekday	0
workingday	0
weathersit	0
temp	0
atemp	0
hum	0
windspeed	0
casual	0
registered	0
cnt	0

## Feature Engineering

```
day <- day_raw %>%
  mutate(date = as_date(dteday),
        year = year(date),
        month = month(date, label = TRUE),
        weekday = wday(date, label = TRUE),
        weekend = if_else(weekday %in% c("Sat", "Sun"), 1, 0)) %>%
  select(date, year, month, weekday, weekend, temp, atemp, hum, windspeed, casual, registered, cnt)
glimpse(day)
```

```
## $ temp      <dbl> 0.3441670, 0.3634780, 0.1963640, 0.2000000, 0.2269570, 0.20~
## $ atemp     <dbl> 0.3636250, 0.3537390, 0.1894050, 0.2121220, 0.2292700, 0.23~
## $ hum       <dbl> 0.805833, 0.696087, 0.437273, 0.590435, 0.436957, 0.518261, ~
## $ windspeed <dbl> 0.1604460, 0.2485390, 0.2483090, 0.1602960, 0.1869000, 0.08~
## $ casual    <dbl> 331, 131, 120, 108, 82, 88, 148, 68, 54, 41, 43, 25, 38, 54~
## $ registered <dbl> 654, 670, 1229, 1454, 1518, 1518, 1362, 891, 768, 1280, 122~
## $ cnt       <dbl> 985, 801, 1349, 1562, 1600, 1606, 1510, 959, 822, 1321, 126~
```

```
day <- day_raw %>%
  mutate(
    dteday = as.Date(dteday),
    year = ifelse(yr == 0, 2011, 2012),
    month = month(dteday, label = TRUE),
    weekday = wday(dteday, label = TRUE),
    weekend = if_else(weekday %in% c("Sat", "Sun"), 1L, 0L)
  ) %>%
  arrange(dteday) %>%
  as_tsibble(index = dteday)

day %>%
  has_gaps()
```

```
## # A tibble: 1 x 1
##   .gaps
##   <lgl>
## 1 FALSE
```

```
summary(day$cnt)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	22	3152	4548	4504	5956	8714

```
stopifnot(nrow(day) == 731) # expected daily records per dataset description
```

## Data Transformation

### Final Dataset Summary

### Data Quality Report

### Data Dictionary

### Summary of Preparation Steps

## Modeling

Modeling Approach

Model Development

Model Evaluation

Model Selection

Summary of Modeling Steps

## Evaluation

Results Summary

Business Impact Assessment

Lessons Learned

Recommendations

## Deployment

Deployment Plan

Monitoring and Maintenance Strategy

Future Work Suggestions

Summary of Deployment Steps

Acknowledgments

About the Author

## Conclusion

Project Summary

Key Takeaways

Final Thoughts



## References

List of References

Citation Style Used

Additional Reading Materials

Acknowledgment of Sources

## Appendix