

BIKE RENTAL



Name: Nidhi Nair
NEU ID: 002371756





OBJECTIVES

- Identify factors influencing rental demand
- Analyze weekday vs. weekend usage

DATASET

The dataset used in this analysis is the "Bike Sharing Dataset" from the UCI Machine Learning Repository. It contains daily aggregated data from Capital Bikeshare in Washington D.C., covering a two-year period from January 1, 2011, to December 31, 2012.

<https://archive.ics.uci.edu/dataset/275/bike+sharing+dataset>

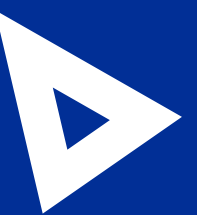


METHODOLOGY



- Data Reading: Load dataset using `read.csv()`
- Data Cleaning: Convert date, converted categorical columns like season, month, and weekday into factors, remove nulls
- Correlation Analysis: Use `corrplot()` to visualize numeric relationships
- Regression Modeling: Build linear model with weather and seasonal factors
- Statistical Testing: T-test for weekday vs weekend rental difference

Data Analysis Process Funnel



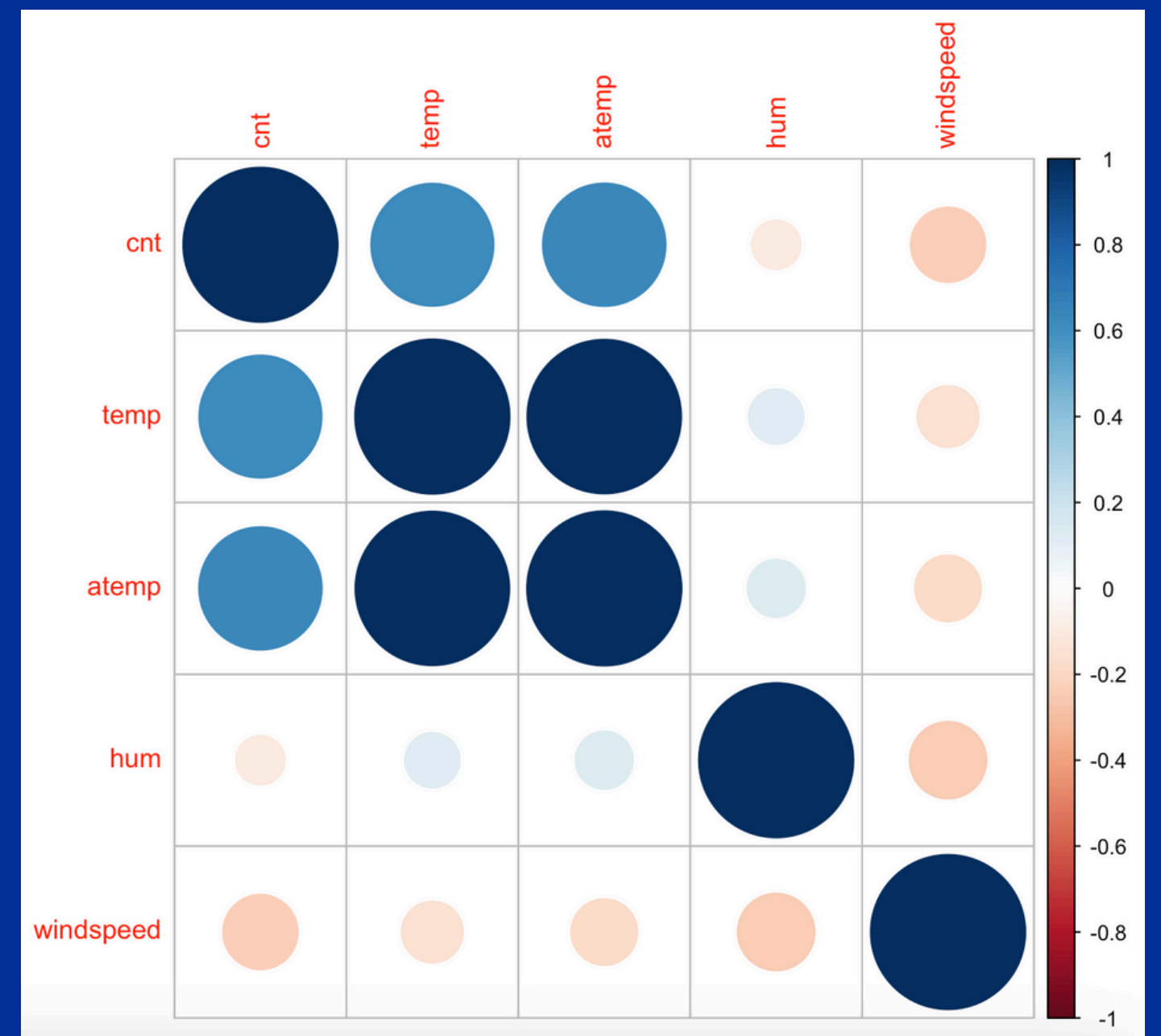


RESULT AND ANALYSIS



```
# Select relevant numeric columns for correlation
numeric_vars <- bike_data %>%
  select(cnt, temp, atemp, hum, windspeed)
# Compute and plot correlation matrix
cor_matrix <- cor(numeric_vars)
corrplot(cor_matrix, method = "circle") # Visualize correlations
```

Analysis: The correlation matrix showed strong positive correlation between cnt (rental count) and temp/atemp, while hum and windspeed showed weak negative correlation.



RESULT AND ANALYSIS

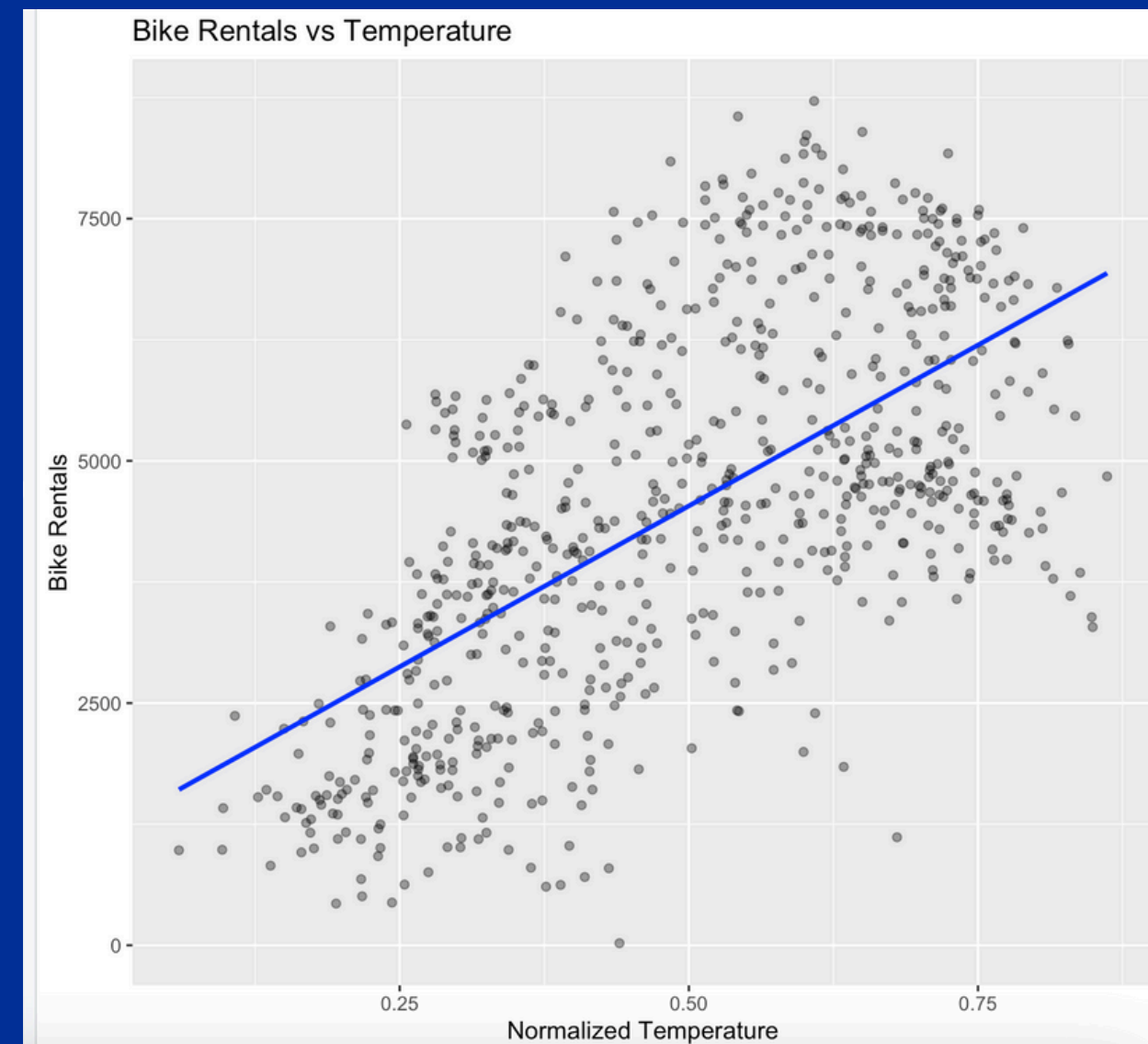
- Temp had a significant positive effect (Estimate = 5652.75, $p < 0.001$)
- hum and windspeed had significant negative effects
- R-squared value: 0.518, indicating that the model explains ~52% of the variation in bike rental counts
- The linear regression model output showed that temperature is a significant predictor. This is further supported by the scatter plot, as temperature increases, we see a clear rise in the number of rentals, confirming our hypothesis.

```
Call:
lm(formula = cnt ~ temp + hum + windspeed + season + weathersit,
    data = bike_data)

Residuals:
    Min       1Q   Median       3Q      Max
-4108.5  -988.9  -198.6   1063.3   4151.1

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3446.01     327.43   10.524 < 2e-16 ***
temp          5652.75     297.26   19.016 < 2e-16 ***
hum          -2359.91     475.79   -4.960 8.79e-07 ***
windspeed    -3358.93     694.73   -4.835 1.63e-06 ***
season         409.64      49.06    8.350 3.46e-16 ***
weathersit    -460.85     120.02   -3.840 0.000134 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1349 on 725 degrees of freedom
Multiple R-squared:  0.5181,    Adjusted R-squared:  0.5148
F-statistic: 155.9 on 5 and 725 DF,  p-value: < 2.2e-16
```





RESULT AND ANALYSIS



- Created a new column `day_type` to classify days as Weekday or Weekend
- Calculated average rentals:
 - Weekday: 4550.57 ; Weekend: 4389.69
- Used a boxplot to visualize the distribution:
 - Wider spread on weekends, but similar medians
- A two-sample t-test was performed to assess if this difference was statistically significant.
- The t-value was 0.9837, with a p-value of 0.3259.
- The 95% confidence interval for the difference ranged from -160.73 to 482.49.
- Since the p-value is greater than the standard threshold of 0.05, we conclude that the observed difference is not statistically significant. This suggests that the slight difference in mean rentals is likely due to random variation, rather than a meaningful trend in the population.



```
99
100 # Run t-test to compare means between weekdays and weekends
101 t_test_result <- t.test(cnt ~ day_type, data = bike_data)
102
103 # Print t-test results
104 print(t_test_result)
```

103:1 # (Untitled) R Script

R 4.4.2 · ~/Work/DSEM/RPrograms/bike_dataset/

```
+ y = "Bike Rentals")
+ `geom_smooth()` using formula = 'y ~ x'
> # Print t-test results
> print(t_test_result)
```

Welch Two Sample t-test

data: cnt by day_type
t = 0.98373, df = 361.39, p-value = 0.3259
alternative hypothesis: true difference in means between group Weekday and group Weekend is not equal to 0
95 percent confidence interval:
-160.7304 482.4914
sample estimates:
mean in group Weekday mean in group Weekend
4550.566 4389.686



CONCLUSION

01

Weather conditions strongly influence daily bike rental demand, with temperature being the most powerful predictor, Humidity and wind speed have significant negative effects on ridership

02

While weekdays show slightly higher average rentals, the difference is not statistically significant.

AND RECOMENDATIONS

01

Develop targeted promotions during periods of expected low demand.

02

Consider covered bike options for harsh weather



THANK YOU

