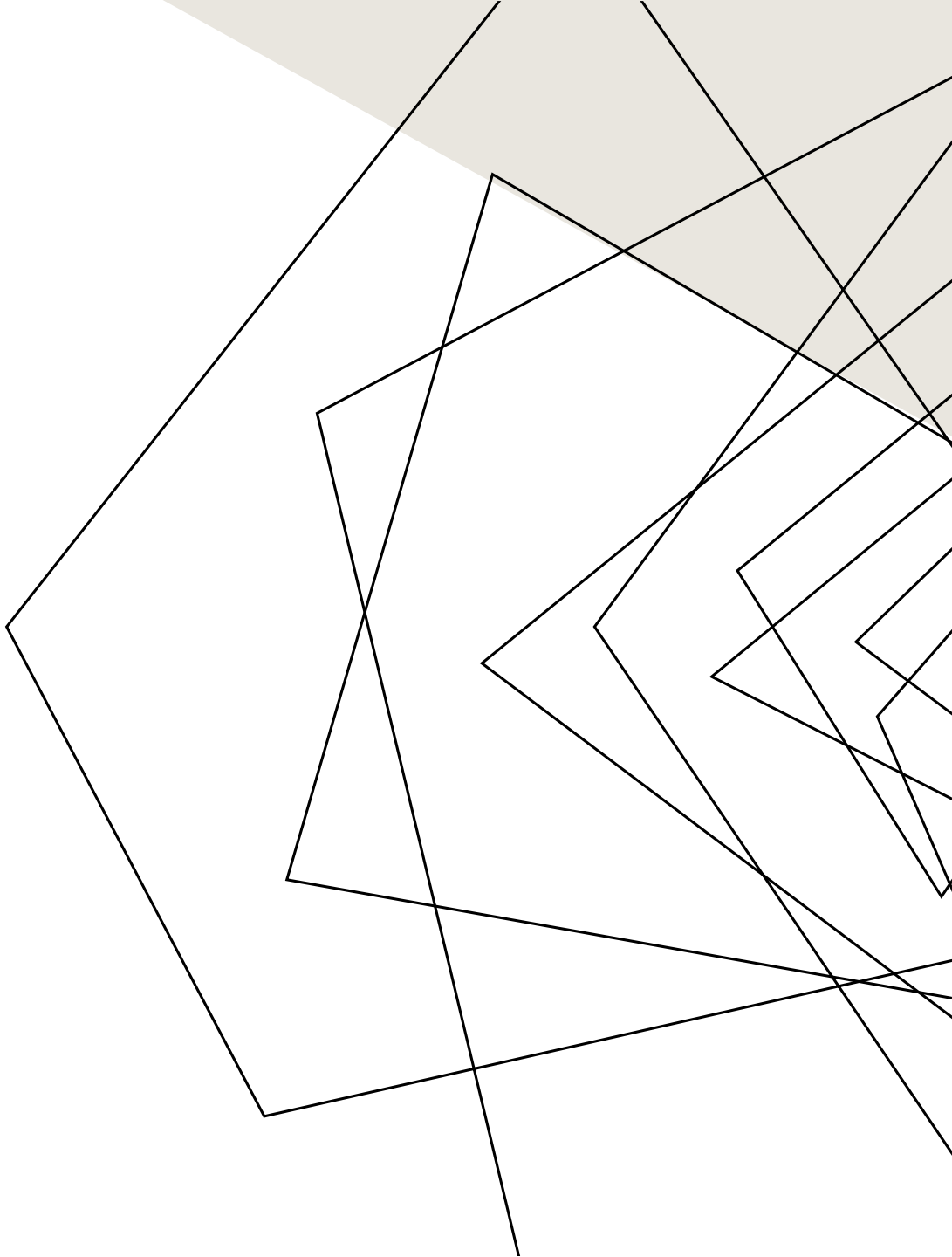
- **PROJECT NAME:** DATA WRANGLING & ANALYSIS
- **ORGANIZATION:** NEXTHIKES IT SOLUTIONS

- **PREPARED BY:** HARSH VARDHAN

INTRODUCTION TO THE DATASET

- **Overview:** The dataset contains bike rental data with weather and time attributes.
- **Purpose:** To clean, preprocess, merge, and analyze data for business insights.
- **Key Features:** Temperature, humidity, windspeed, working days, rentals.
- **Goal:** Identify trends, detect anomalies, and generate insights for decision-making.
- **Visuals:** Infographic showing data acquisition, wrangling, and analysis steps.



DATASET 1 - INITIAL INSIGHTS

- **Description:** Dataset 1 records hourly bike rentals with weather details.
- **Key Attributes:** `temp`, `humidity`, `windspeed`, `cnt`
- **Main Plots:**
  - **Temperature vs. Rentals:** Shows rental trends based on temperature.
  - **Temperature Distribution:** Identifies temperature variations in the dataset.
- **Insights:**
  - More rentals occur at moderate temperatures.
  - High humidity reduces rentals, indicating weather sensitivity.
- **Visuals:** Well-designed line and histogram charts.

DATASET 1 – DATA CLEANING & PREPROCESSING

- **Steps Taken:**
  - Checked for missing values (Handled using median imputation).
  - Removed duplicate records (if any).
  - Converted `dteday` to datetime format.
- **Outliers:**
  - Boxplot revealed extreme values in `casual` and `registered` columns.
  - Outliers were handled using the **IQR method**.
- **Visuals:**
  - Boxplot for `cnt` rentals.
  - Summary statistics table.

DATASET 2 – ANALYSIS & FINDINGS

- **Description:** Contains additional details like registered vs casual rentals.
- **Key Attributes:** `atemp`, `casual`, `registered`
- **Main Plots:**
    - **Registered vs. Casual Users:** Shows proportion of users.
    - **Windspeed vs. Rentals:** Tests correlation between windspeed and rentals.
- **Insights:**
    - **Registered users dominate total rentals** (higher customer retention).
    - Windspeed has minimal effect on rental count.
- **Visuals:** Pie chart for user comparison, scatter plot for windspeed impact.

DATASET 2 – DATA CLEANING & PREPROCESSING

- **Steps Taken:**
  - Checked missing values and handled them.
  - Removed unnecessary columns (`Unnamed: 0`).
  - Converted date format for consistency.
- **Outliers:**
  - Identified extreme values in `casual` and `registered` columns.
  - Log transformation applied for normalizing data.
- **Visuals:** Heatmap showing data correlations.

DATASET 3 – HOURLY RENTALS & WEATHER IMPACT

- **Description:** This dataset includes hourly rental details with weather impact.
- **Key Attributes:** `season`, `weathersit`, `hr`
- **Main Plots:**
    - **Hourly Rental Trends:** Rentals peak during commuting hours.
    - **Weather Condition Impact:** Fewer rentals in poor weather.
- **Insights:**
    - Highest rentals observed in the morning and evening.
    - Severe weather conditions reduce rentals drastically.
- **Visuals:** Line graph for hourly trends, bar chart for weather impact.

DATASET 3 – DATA CLEANING & PREPROCESSING

- **Steps Taken:**
  - Handled missing values using interpolation.
  - Verified data integrity (no duplicate timestamps).
  - Converted categorical variables (seasons) into numerical format.
- **Outliers:**
  - Some extreme high rental values removed after analysis.
- **Visuals:** Boxplot for season-wise rentals.

MERGING DATASET 1 & 2 – COMBINED INSIGHTS

Process:

- Merged on `instant` column.
- Unified rental data with user details.

- **Main Plots After Merging:**
  - **Temperature vs. Rentals (Combined Data).**
  - **Registered vs. Casual Users (Combined Data).**
- **Key Insights:**
  - Improved data consis** Merging Dataset 1 & 2 - Combined Insights**
    - tency after merging.
  - More accurate rental trends.
- **Visuals:** Multi-line graph & scatter plot for rentals vs temp.

FINAL MERGED DATASET (1.1 + 3)

- **DESCRIPTION:** INTEGRATED ALL DATASETS FOR FINAL INSIGHTS.
- **MAIN VISUALS:**
- **Final Rental Distribution:** Histogram of total rentals.
- **Seasonal Impact:** Rentals across different seasons.
- **INSIGHTS:**
- Skewness and extreme values corrected.
- Strong seasonality pattern detected.
- **VISUALS:** BAR GRAPH FOR SEASONAL RENTALS.

OUTLIER HANDLING

- **Identified Columns:** `casual`, `registered`, `cnt`
- **Method Used:** Interquartile Range (IQR) to remove extreme values.

- **Justification:**
  - Improved model accuracy by eliminating data distortions.
  - Outliers removed in business-relevant way.
- **Visuals:** Before vs After Boxplots.

CORRELATION ANALYSIS

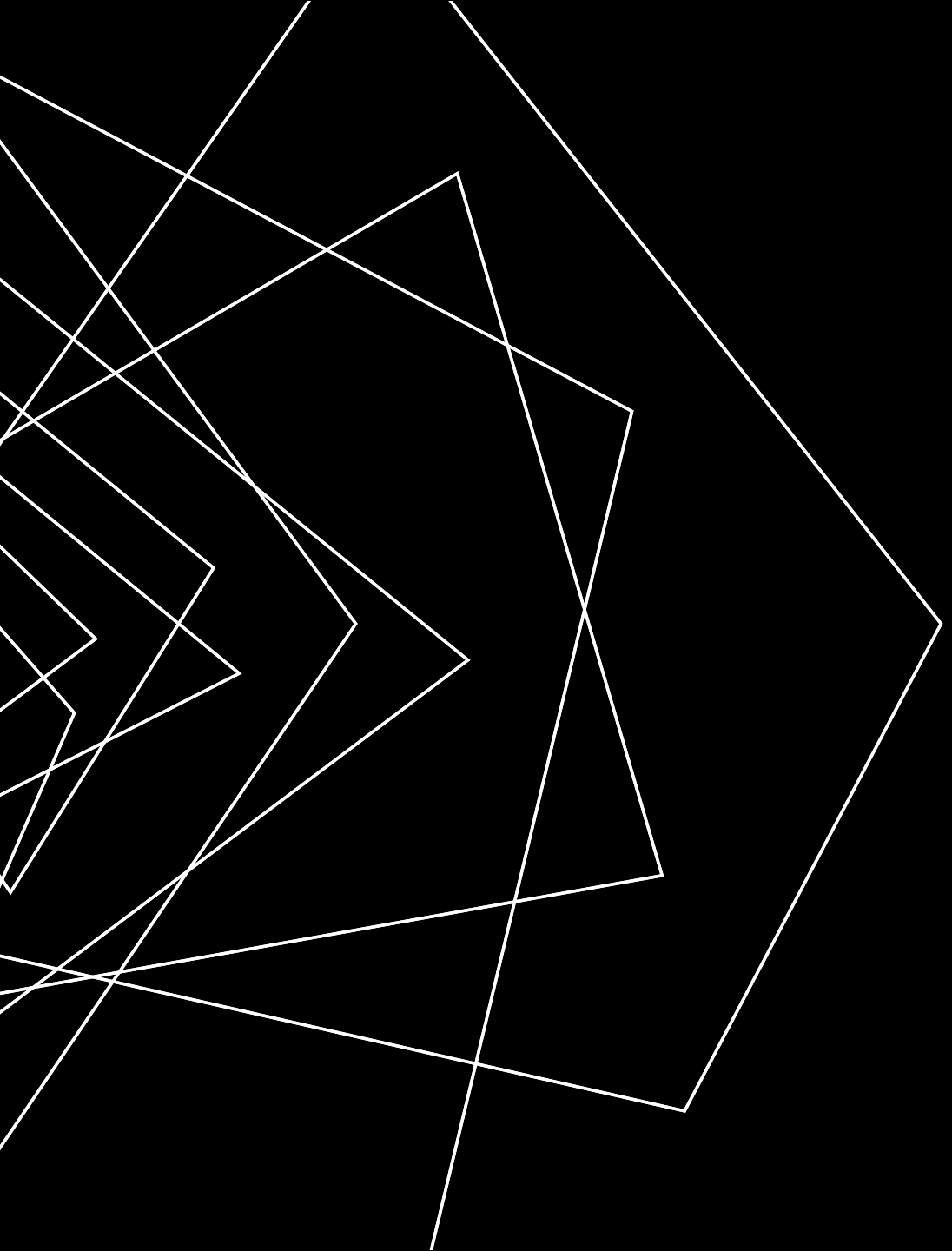
- **Key Findings:**
  - **Positive Correlation:** `temp` & `cnt`, `registered` & `cnt`.
  - **Negative Correlation:** `humidity` & `cnt`.
- **Impact:**
  - Temperature affects rentals strongly.
  - **Visuals:** Heatmap of correlations.
  - More registered users indicate higher long-term retention.

SKEWNESS & KURTOSIS

- **SKEWNESS ANALYSIS:**
- `casual` and `registered` are highly skewed.
- **KURTOSIS INSIGHTS:**
- Presence of extreme values in the dataset.
- **HOW WE HANDLED IT:**
- Applied **log transformation** to reduce skewness.
- **VISUALS:** BEFORE VS AFTER DISTRIBUTION GRAPHS.

CONCLUSION & FUTURE RECOMMENDATIONS

- **Final Observations:**
    - Rentals depend highly on temperature and working hours.
    - Weather conditions have a strong impact on demand.
- **Recommendations:**
    - Implement dynamic pricing based on peak rental times.
    - Improve bike availability based on hourly demand.
- **Visuals:** Summary infographic.



THANK YOU

"Thank you for your time and attention! Looking forward to your feedback and discussion."