EasyChair Preprint

№ 15641

Public and Private Transportation: Which Do People Prefer and Why?

Saikal Batyrbekova and Gulnaz Gimaletdinova

January 6, 2025

RESEARCH PAPER ON PUBLIC AND PRIVATE TRANSPORTATION: Which Do People Prefer and Why?

Statistics and Probability Course - Fall 2024 Semester

Batyrbekova Saikal *- saikal.batyrbekova@alatoo.edu.kg*

Gimaletdinova Gulnaz *- gulnaz.gimaletdinova@alatoo.edu.kg*

Ala-Too International University

ABSTRACT

This research explores the transportation preferences of Bishkek residents and the factors influencing their choices. Data from 74 participants was analyzed to understand how often public and private transport is used, the costs involved, and travel distances. Using Python, I identified patterns and relationships, such as the impact of cost and convenience on decisions. The results showed that public transport is popular due to its affordability, while private transport is preferred for its comfort and flexibility. Visual tools like histograms, bar charts, and scatter plots were used to present the findings clearly. This study sheds light on people's transportation habits, providing ideas for future improvements.

Keywords: Public transportation, private transportation, transportation choices, frequency, cost, distance, data analysis, patterns, trends, convenience, graphs and charts, transportation options.

1. INTRODUCTION

Transport plays a huge role in everyday life, and in Bishkek, how people get around is influenced by things like cost, convenience, and available options. In recent years, there have been some improvements to the city's transportation system, but there are still some challenges. Public transport, especially buses, is an important option for many people, particularly for school students and university students. One of the key changes has been the introduction of the Tulpar card, which offers discounted bus fares for students, disabled people, and pensioners. This helps make transportation more affordable for these groups.

Since 2021, Bishkek has also added new buses to the public transport system. These buses are meant to be more comfortable, efficient, and eco-friendly, which is great for improving the overall experience. However, there are still issues, such as overcrowding during rush hours and a need for more routes to serve the growing population.

Taxis are also widely used in Bishkek, with services like Yandex and Navi being the most popular. These companies charge based on distance, time of day, and demand, but taxi fares can be higher than bus fares, which makes them less affordable for people who rely on public transport for daily commuting.

For this project, I wanted to research what people in Bishkek prefer when it comes to transports and which they use more often. I collected data through a Google Forms survey from 74 people, asking about their transport habits, how much they spend, and what kind of transport they prefer. The goal is to understand which type of transport people in Bishkek use the most and what their preferences are.

This study is important because Bishkek is growing, and understanding what kind of transportation people prefer can help us better understand the needs of the city's residents. This paper will explain how I collected and analyzed the data, present the findings, and discuss which modes of transportation are more popular in Bishkek.

As part of this project, data was analyzed to which artificial intelligence (AI) will be applied in the future, which shows good results not only in character recognition, but also finds successful application in other areas.

---

2. METHODS

In this research, I wanted to explore how people in Bishkek prefer to travel. I collected data through a survey with 74 participants, and to analyze this data, I used different statistical methods to help understand any patterns or connections. Here's techniques I used:

Descriptive Statistics

To get a general overview of the data, I started by calculating basic descriptive statistics. These are simple ways to summarize the information and understand how the participants are responding to the questions. The main measures I used were:

Mean (Average):

The mean was used to calculate the average answer for questions like, "On average, how many days per week do you use public transport?" This helps to see how often participants are using public transport on average. Why did I use a bar graph? A bar graph is

great for showing **quantitative** data, like how many days people use public transport. It helps you easily compare how often people use it and see which answer is most common.

Median (Middle Value):

The median was used to find the middle value for questions about commute time, like, "How many minutes do you spend commuting each day?" The median is helpful because it tells us the middle value, so it's not affected by really long or short commute times. This gives a better idea of the typical commute time for most people. Why a line graph? A line graph is best for **continuous data** like how much time people spend commuting. It helps show patterns and trends, like if most people have similar commute times.

Mode (Most Common Answer):

The mode was used to find the most common response for questions like, "Which type of transportation do you use the most regularly?" The mode tells us which transportation option (bus, taxi, etc.) is most popular among participants. Why a bar graph? Again, a bar graph works best here because it allows you to see which transportation type is used most often. Since the responses are categories (e.g., bus, walking, taxi), a bar graph helps show which one stands out as the most frequently chosen option.

In conclusion, using descriptive stats and graphs helped me analyze the data clearly. The bar graph showed how often people use public transport and which options are most popular. The line graph tracked commute times, revealing common trends. The mode helped highlight the most frequent responses. The results will be shown in the next pages.

Distribution Analysis

To perform a distribution analysis of transportation expenses, I included the question *"On average, how much do you spend on transportation per day?"* in my survey. The aim of this question was to understand the daily spending habits of respondents on transportation. By analyzing the distribution of these expenses, I can gain insights into the general patterns of

transportation spending and better understand which modes of transport are most commonly used and how much they cost on average.

To analyze the distribution, I used a **histogram** and a **Q-Q plot**.

- The **histogram** was used to visually represent the frequency of different spending amounts, allowing me to see how many people fall into various spending categories. This helped to identify whether most people spend within a certain range, or if there are outliers.
- The **Q-Q plot** was used to check the normality of the data. By comparing the distribution of the data to a normal distribution, I could determine if the spending data follows a bell curve, or if it shows any signs of skewness or irregular patterns.

These tools were chosen because they provide both a visual and statistical way to understand the distribution of the data, helping to draw conclusions about transportation spending trends.

---

3. RESULTS

Descriptive Statistics Graphs

Visualization of the Mean: Days of public transport use per week

To analyze the average number of days respondents use public transportation, I calculated the **Mean** using the data collected with the question: *How many days a week do you use public transportation?* The responses ranged from 1 to 7 days per week, with the corresponding number of respondents as follows:

Days per Week	1 day	2 days	3 days	4 days	5 days	6 days	7 days
Responders	7	1	5	5	22	16	18

The mean was calculated using the formula:

$$\text{Mean} = \frac{\sum(x_i \cdot f_i)}{N}$$

Where:

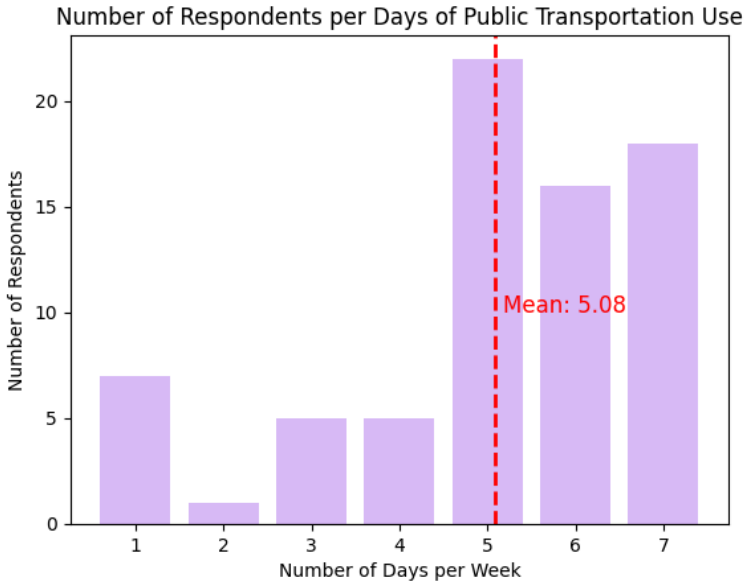
- Xi = each number of days (1, 2, 3, etc.)
- Fi = number of respondents for each corresponding day
- N = total number of respondents

Substituting the values:

$$\text{Mean} = \frac{(1 \times 7) + (2 \times 1) + (3 \times 5) + (4 \times 5) + (5 \times 22) + (6 \times 16) + (7 \times 18)}{74}$$

$$\text{Mean} = \frac{376}{74} \approx 5.08$$

Thus, the average number of days respondents use public transportation is approximately 5.08 days per week. I will use a bar graph to visualize the frequency of each response, and annotate the **Mean** on the chart:



Python code snippet that calculated and visualized the Mean:

```
import matplotlib.pyplot as plt

import numpy as np

# Data

days = [1, 2, 3, 4, 5, 6, 7]

responders = [7, 1, 5, 5, 22, 16, 18]

# Calculate the mean

total_responders = sum(responders)

weighted_sum = sum([days[i] * responders[i] for i in range(len(days))])

mean = weighted_sum / total_responders

# Plotting the bar chart with lighter purple color

plt.bar(days, responders, color='#D8B9F6')  # Lighter shade of purple

# Adding labels and title

plt.xlabel('Number of Days per Week')

plt.ylabel('Number of Respondents')

plt.title('Number of Respondents per Days of Public Transportation Use')

# Annotate the mean

plt.axvline(mean, color='red', linestyle='dashed', linewidth=2)

plt.text(mean + 0.1, 10, f'Mean: {mean:.2f}', color='red', fontsize=12)

# Show plot (This line won't run in Google Doc)

#plt.show()

# Print the mean

#print(f"The mean number of days per week respondents use public
transportation is: {mean:.2f}")
```

Visualization of the Median: Time spent commuting each day

To analyze the central tendency of the time spent commuting each day, I calculated the **Median** using the data collected with the question: *How many minutes do you spend commuting each day?* The responses ranged from 5 minutes to 2 hours, with the corresponding number of respondents as follows:

Time Spent Commuting	5-10 min	15 min	30 min	45 min	1 hour	1.5 hour	2 hours
Responders	7	10	9	10	14	9	15

The median was calculated using the formula:

$$\text{Median} = L + \left(\frac{\frac{N}{2} - F}{f} \right) \times h$$

Where:

- **L** = 45 (lower boundary of the median class: 45 minutes)
- **N** = 74 (total number of respondents)
- **F** = 26 (cumulative frequency up to 30 minutes)
- **f** = 10 (frequency of the 45 minutes class)
- **h** = 15 (class width: 45 minutes - 30 minutes)

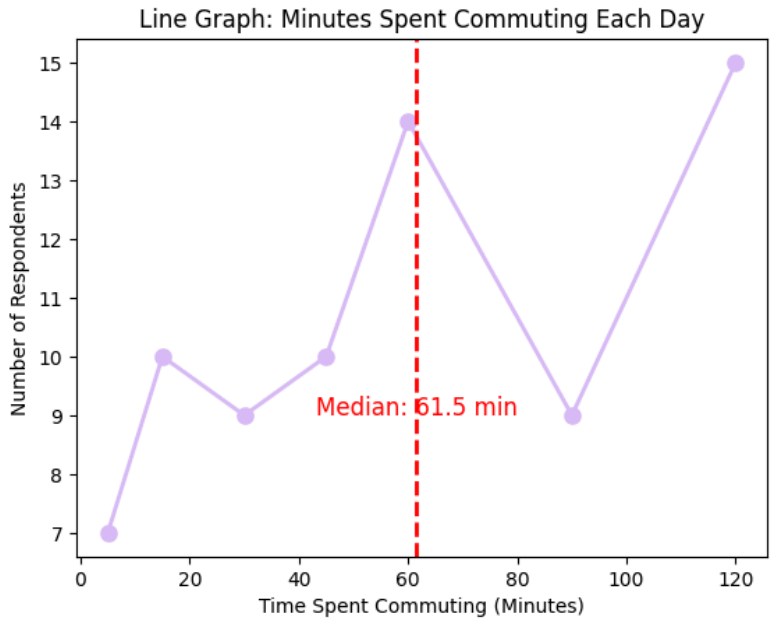
Substituting the values:

$$\text{Median} = 45 + \left(\frac{\frac{74}{2} - 26}{10} \right) \times 15$$

$$\text{Median} = 45 + \left(\frac{37 - 26}{10} \right) \times 15$$

$$\text{Median} = 45 + \left(\frac{11}{10} \right) \times 15 = 45 + 16.5 = 61.5 \text{ minutes}$$

Thus, the median number of minutes respondents spend commuting each day is approximately 61.5 minutes. I will use a line graph to visualize the frequency of each response, and annotate the **Median** on the chart:



Python code snippet that calculated and visualized the Median:

```
import matplotlib.pyplot as plt

import numpy as np

# Data for minutes spent commuting

commuting_times = [5, 15, 30, 45, 60, 90, 120]  # Time in minutes

responders = [7, 10, 9, 10, 14, 9, 15]          # Number of respondents

# Plotting the line graph
```

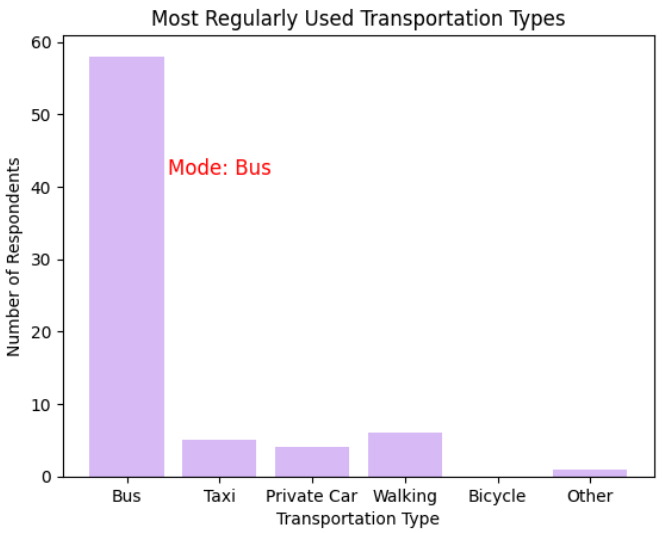
```
plt.plot(commuting_times, responders, marker='o', color='purple', linestyle='-',
linewidth=2, markersize=8)

# Adding labels and title

plt.xlabel('Time Spent Commuting (Minutes)')

plt.ylabel('Number of Respondents')

plt.title('Line Graph: Minutes Spent Commuting Each Day')

# Annotating the median

median = 61.5

plt.axvline(median, color='red', linestyle='dashed', linewidth=2)

plt.text(median + 2, 3, f'Median: {median} min', color='red', fontsize=12)

# Show plot

plt.show()

# Print the median

print(f"The median number of minutes spent commuting each day is: {median}
minutes")
```

Visualization of the Mode: Days of public transport use per week

To analyze the central tendency of the most regularly used transportation, I calculated the **Mode** using the data collected with the question: *Which type of transportation do you use the most regularly?* The responses ranged from "Bus" to "Bicycle," with the corresponding number of respondents as follows:

Transport Type	Bus	Taxi	Private Car	Walking	Bicycle	Other
Responders	58	5	4	6	0	1

Thus, the **Mode** of transportation used most regularly by respondents is **Bus**, which was selected by 58 individuals. I will use a bar chart to visualize the frequency of each response, and annotate the **Mode** on the chart:



Python code snippet that calculated and visualized the Mode:

```
import matplotlib.pyplot as plt

# Data for most regularly used transportation

transportation_types = ['Bus', 'Taxi', 'Private Car', 'Walking', 'Bicycle',
'Other']

responders = [58, 5, 4, 6, 0, 1]

# Plotting the bar chart

plt.bar(transportation_types, responders, color='#D8B9F6')

# Adding labels and title

plt.xlabel('Transportation Type')

plt.ylabel('Number of Respondents')

plt.title('Most Regularly Used Transportation Types')

# Annotating the Mode with adjusted position

mode = 'Bus'  # The mode is the transportation used by the most respondents

plt.text(1, 40 + 1, f'Mode: {mode}', color='red', fontsize=12, ha='center',
va='bottom')

# Show plot

plt.show()
```

```
# Print the Mode

print(f"The mode of transportation used most regularly is: {mode}")
```

Distribution Analysis

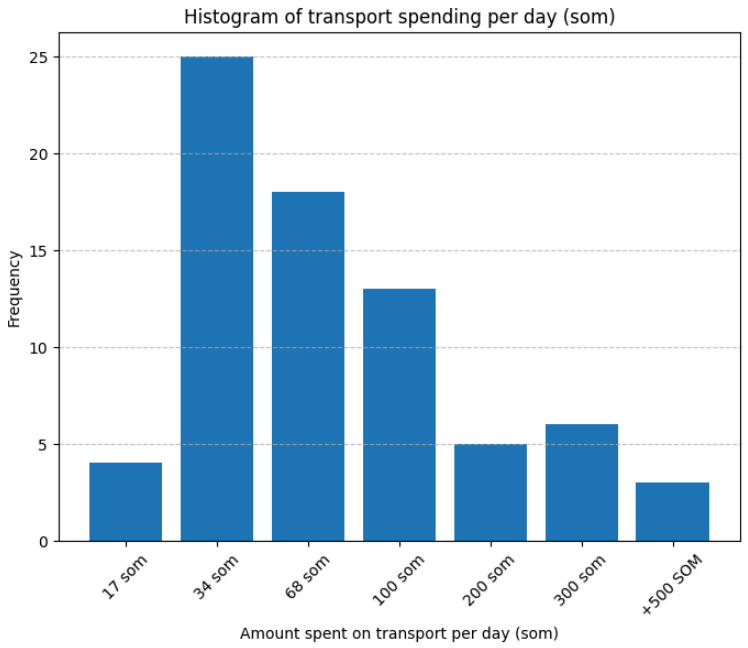
Analysis of Transport Spending Distribution: *On average, how much do you spend on transport per day?*

Firstly, I need to organize the data I collected:

Amount (SOM)	**Frequency (Responders)**
17	4
34	25
68	18
100	13
200	5
300	6
500+	3
	Total: 74

1. Histogram

To better understand how people spend on transport daily, I first created a histogram. A histogram is a type of bar chart that shows how many people fall into different spending ranges.



The histogram shows the distribution of daily transport spending among 74 responders. It illustrates how many people spend different amounts on transportation each day.

What we can observe:

- **Peak Spending:** The highest number of people (25) spend around 34 som per day, meaning a lot of people spend that much on transport each day. This is probably because public transport like buses and marshrutkas is pretty cheap.
- **Some people spend more:** There are smaller bars for higher amounts, like 68, 100, and even 200+ som. This could be because they live far away and make multiple transfers.
- **Right Skew:** The distribution is skewed to the right, meaning there are a few people who spend significantly more than the majority.
- **Range of Spending:** The spending ranges from a minimum of 17 som to over 500 som.

This histogram provides a visual representation of the variability in daily transport expenses among the respondents.

Code Snippet:

```
import pandas as pd import matplotlib.pyplot as plt
```

```
# Create a dictionary from the data

data = {

   "17 som": 4, "34 som": 25, "68 som": 18, "100 som": 13,

   "200 som": 5, "300 som": 6,  "+500 SOM": 3}

# Create a DataFrame from the dictionary

df = pd.DataFrame.from_dict(data, orient='index', columns=['Count'])

plt.figure(figsize=(8, 6)) # Create a histogram

plt.bar(df.index, df['Count'])

plt.xlabel('Amount spent on transport per day (som)')

plt.ylabel('Frequency')

plt.title('Histogram of transport spending per day (som)')

plt.xticks(rotation=45)  # Rotate x-axis labels for better readability

plt.grid(axis='y', linestyle='--', alpha=0.7)

plt.show()
```

Conclusion

The lower spending values could indicate that people are prioritizing cheaper public transport options. As the spending increases, it's possible that those individuals are using taxis or personal vehicles, which are more expensive but offer more convenience.

---

2. Q-Q Plot (Quantile-Quantile)

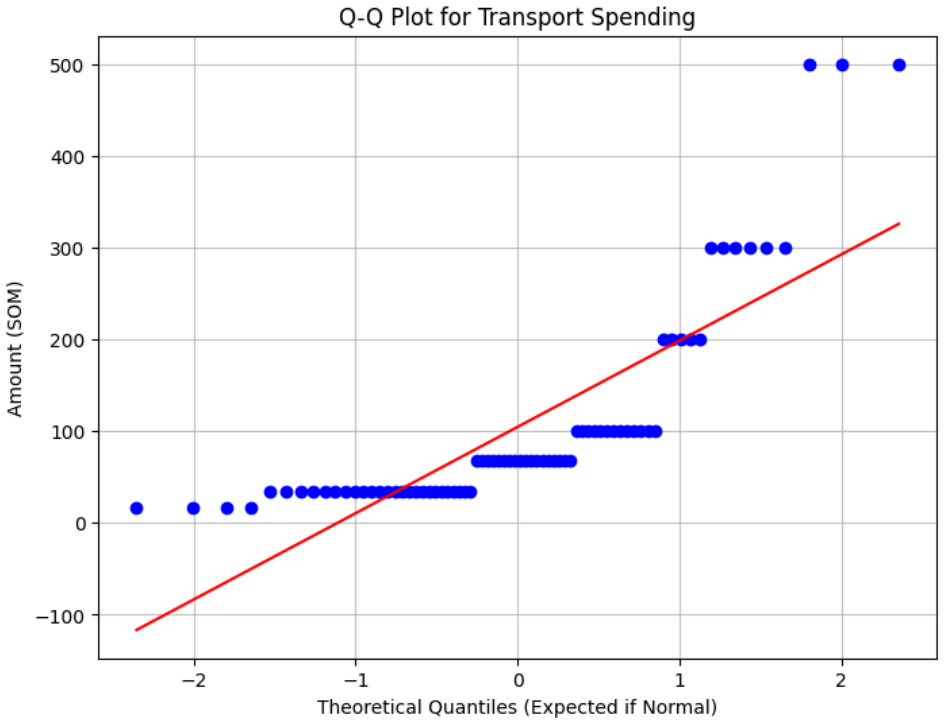
Secondly, a Q-Q plot helps us check if data follows a normal distribution, that's why I will use Q-Q plot. A normal distribution is what people often call a "bell curve" — it's symmetric, with most values clustering around the average, and fewer values at the extremes.

In a Q-Q plot:

- **Red Line**: This shows what the data **should look like** if it were perfectly normal.
- **Blue Dots**: These represent the **actual data points** from your dataset.

If the blue dots align closely with the red line, my data is probably normal. If they don't, it's not normal. I use the Q-Q plot because it's a quick way to visually check for normality. It can show patterns like skewness (data leaning heavily to one side) or outliers (values that are way off from the rest).

Here you can see my Q-Q plot:



What does the Q-Q Plot show for transport spending?

At the Ends (Tails):

- The blue dots curve away from the red line, especially at the top (right tail).
- This happens because a small number of people spend a lot more (like 200 som, 300 som, or 500+ som) compared to most others. This creates a right skew in the data.
- On the low end, there's also a bit of a curve because of people spending very little (like 17 som).

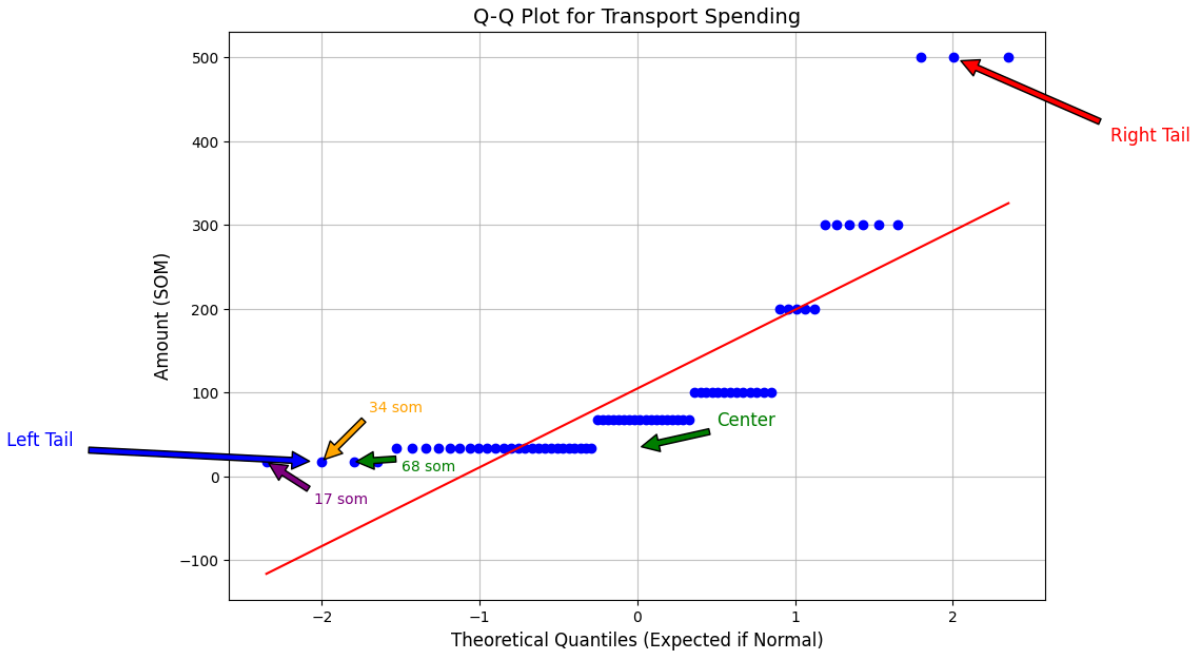
In the Middle:

- The blue dots around the middle mostly stay near the red line, but they're not perfectly aligned. This shows the central part of the data (around 34 som) clusters more tightly than it would in a normal distribution.

X and Y Axes:

- The X-axis shows where the data points should be if the data were normal.
- The Y-axis shows the amount of SOM

More detailed Q-Q plot:



Code Snippet:

```
import numpy as np

import scipy.stats as stats

import matplotlib.pyplot as plt

# Original grouped data

spending_values = [17, 34, 68, 100, 200, 300, 500]

frequencies = [4, 25, 18, 13, 5, 6, 3]

# Expand the grouped data into individual data points
```

```
data = np.repeat(spending_values, frequencies)

# Generate the Q-Q plot

plt.figure(figsize=(8, 6))

stats.probplot(data, dist="norm", plot=plt)  # dist="norm" ensures comparison with
a normal distribution

plt.title("Q-Q Plot for Transport Spending")

plt.xlabel("Theoretical Quantiles (Expected if Normal)")

plt.ylabel("Amount (SOM)")

plt.grid(alpha=0.7)

plt.show()
```

How do we know it's not normal: The blue dots don't follow the red line.

At the top end, they pull away because a few people spend way more than average.

- At the bottom, they pull away slightly because of clustering at lower values.
- This mismatch, especially at the ends, means the data isn't symmetric or "bell-shaped."

Conclusion

Based on the Distribution Analysis, including the Q-Q plot and histogram, I can conclude that the daily transport spending of the respondents shows a clear preference for public transport. Here's why:

1. Most people spend around 34 som: The majority of respondents (25 people) reported spending around 34 som per day. This amount is typical for affordable public transport, like buses and marshrutkas, which are commonly used in Bishkek.

2. Right skewness in spending: While some respondents report spending significantly more (like 200+ som), the overall pattern is right-skewed, meaning most people spend relatively little on transport. This suggests that public transport is a convenient and

affordable option for most people.

3. Affordability of public transport: Given that the bulk of people are spending less, we can infer that public transport remains the most affordable choice for most individuals, making it a practical option for daily commuting.

In conclusion, the data shows that **the majority of people prefer public transport** due to its affordability, as evidenced by the relatively low daily transport spending reported by most respondents.

---

4. DISCUSSION

The findings of this research reveal notable trends in transportation preferences among Bishkek residents, highlighting public transport as the dominant mode due to its affordability, while private transport is often chosen for comfort and flexibility. These insights stem from the analysis of data collected from 74 respondents, who shared information about their transportation habits, costs, and distances traveled.

Interpretation of Results

The data showed that the majority of participants use public transportation approximately five days a week, spending an average of 34 som per day. This aligns with the perception that buses and marshrutkas are affordable and widely accessible options for commuting. The histogram analysis further emphasized this preference, revealing that most respondents fall within a low spending range, reflective of public transport fares.

However, the data also pointed to a subset of individuals who spend significantly more, likely opting for taxis or private vehicles. The right-skewed distribution and Q-Q plot analysis confirmed this variation, indicating that while public transport is predominant, private transport remains a choice for those seeking convenience or living in areas not well-served by public routes.

Visualization through bar graphs, histograms, and Q-Q plots proved instrumental in identifying these patterns. The mode of transport used—public buses—was the most

frequently selected, underscoring the accessibility and cost-effectiveness of the public transport system.

Possible Limitations

Despite the valuable insights gained, this study faced certain limitations:

1. *Sample Size:* The survey included only 74 participants, which may not fully represent the diverse demographics of Bishkek. A larger sample size could provide more comprehensive insights.
2. *Limited Variables:* The survey focused primarily on cost, frequency, and distance. Other influential factors, such as safety, cleanliness, and availability of alternative routes, were not considered.
3. *Self-Reported Data:* Participants provided self-reported information, which might include inaccuracies or biases, particularly regarding spending habits or preferred modes of transport.
4. *Focus on Younger Demographics:* The study predominantly surveyed school and university students, which might skew results toward their preferences and neglect the habits of older age groups.

Recommendations

To build on these findings and address the identified limitations, the following recommendations are proposed:

1. *Expand the Survey:* Future studies should include a larger and more diverse sample, encompassing different age groups, socioeconomic backgrounds, and geographic areas within Bishkek.
2. *Broaden the Scope*: Incorporate additional variables such as travel time reliability, environmental impact, and the role of digital tools like TULPAR cards in improving access.
3. *Focus on Accessibility:* Address the gaps in public transport routes and prioritize underserved areas to make public transportation a viable option for more residents.
4. *Enhance Public Transport:* Continue investments in eco-friendly and efficient buses, while addressing challenges like overcrowding during peak hours.

5. *Encourage Data-Driven Policies:* Use the insights from similar studies to develop data-driven policies aimed at balancing affordability and convenience in the city's transportation system.

---

5. CONCLUSION

This research was honestly a mix of tough and fun! At first, figuring out how to organize and analyze the data felt overwhelming, but as I got into it, I actually started to enjoy the process. It gave me a new perspective on how transportation works in Bishkek. The study showed just how important public transport is for people's daily lives. If the issues we found—like overcrowding and limited routes—are fixed, the city could have a way better system that works for everyone.

As someone who uses public transport almost every day, I can say it's pretty convenient and affordable for me. That said, rush hours and packed buses are definitely not my favorite parts. This project made me realize just how much we rely on public transport and why improving it could make a big difference for a lot of people, especially students like me.

---

6. REFERENCES

- *BeepKart. (n.d.). Public transport vs private transport. BeepKart. Retrieved from https://beepkart.com/blog/public-transport-vs-private-transport*
- *Dorduz, J. (n.d.). Language models visualization. Retrieved from https://juanitorduz.github.io/lm_viz/*
- *International Marketing Association. (n.d.). Distribution analysis. Retrieved from https://www.intemarketing.org/marketing-information/marketing-analysis/distribution-analysis*
- Shaiakhmetov, Dim, et al. "Criteria for Assessing the Quality of Educational Games and Their Impact on Student Learning Outcomes." *Proceedings of the International Conference on Computer Systems and Technologies 2024*. 2024.

- Shaiakhmetov, Dim, et al. "Morphological classification of galaxies using SpinalNet." *2021 16th International Conference on Electronics Computer and Computation (ICECCO)*. IEEE, 2021.

- Toktosunova, Akbiike, et al. "Developing an Artificial Intelligence Tool for Image Generation Using a Unique Dataset with Image-to-Image Functionality." *Proceedings of the International Conference on Computer Systems and Technologies 2024*. 2024.

- Sadriddin, Zuhra, Remudin Reshid Mekuria, and Mekia Shigute Gaso. "Machine Learning Models for Advanced Air Quality Prediction." *Proceedings of the International Conference on Computer Systems and Technologies 2024*. 2024.

- Gaso, Mekia Shigute, et al. "Utilizing Machine and Deep Learning Techniques for Predicting Re-admission Cases in Diabetes Patients." *Proceedings of the International Conference on Computer Systems and Technologies 2024*. 2024.

- Comparison and Selection of an Edge Detection Filters for ECG Applications
  Dim Shaiakhmetov, Mohd. Tauheed Khan, Gulnaz Gimaletdinova
  Alatoo Academic Studies 23 (1), 509 - 519