Project Pitch: *Sarajevo Apartment Price prediction*

Emre Arapčić-Uevak Mustafa Sinanovic

International University of Sarajevo
Faculty of Engineering and Natural Sciences

November 11, 2025

What will be covered?

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

What is the real-world problem?

In the modern housing market, both **apartment sale prices** and **rental prices** can vary drastically depending on factors such as location, size, number of rooms, building condition, and surrounding amenities.

What is the real-world problem?

However, in many listings, especially on local platforms, price information is often missing, incomplete, or inconsistent. This makes it difficult for potential buyers, tenants, and even property owners to assess the true value of an apartment or to set a fair and competitive price.

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

Why is it important or relevant?

For most people, housing is the largest financial decision they will make, yet reliable pricing information is often unavailable.

When price data is missing or unrealistic, users have no objective way to determine whether a listing is underpriced, overpriced, or fair.

Why is it important or relevant?

A system capable of accurately predicting housing and rental prices would:

- Help buyers and renters make more informed decisions.
- Allow property owners to estimate realistic values for their listings.
- Enable market analysis, regional comparisons, and price trend visualizations for researchers or real estate agencies.

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

Our Goal

Our goal is to build a predictive model that can estimate both apartment sale prices and rental prices based on publicly *available data*, and integrate the model into an application for ease of use.

Our Goal

Our goal is to build a predictive model that can estimate both apartment sale prices and rental prices based on publicly *available data*, and integrate the model into an application for ease of use.

Some concrete examples:

1. Interactive map

Our Goal

Our goal is to build a predictive model that can estimate both apartment sale prices and rental prices based on publicly *available data*, and integrate the model into an application for ease of use.

Some concrete examples:

1. Interactive map
2. Price Fairness Indicator

Our Goal

Our goal is to build a predictive model that can estimate both apartment sale prices and rental prices based on publicly *available data*, and integrate the model into an application for ease of use.

Some concrete examples:

1. Interactive map
2. Price Fairness Indicator
3. Price Prediction Tool for Listings without given prices

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

What has already been explored

The only work we were able to find was Real Estate Market in the Federation of B&H for Year 2022 by FGU (Foundation for Governance and Urban Development).

What has already been explored

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

What do we offer that is new?

The only existing reports that we were able to find are statistical analyses published by organizations like FGU, which provide general information about the market but doesn't offer predictive capabilities.

What do we offer that is new?

To the best of our knowledge, there is no publicly available tool or application that provides apartment price predictions specifically for Sarajevo or Bosnia and Herzegovina.

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

Collecting The Data

We plan to collect data from popular real estate listing websites in Bosnia and Herzegovina, such as:

- OLX.ba
- Nekretnine.ba
- HaloOglasi.com

via web scraping.

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

Overview of Features

We extracted the following features from the real estate listings:

- **Price**
- **Municipality**
- **Condition**
- **Ad Type**
- **Property Type**
- **Number of Rooms**
- **Area (m^2)**
- **Furnishing Level**
- **Floor Level**
- **Heating Type**

Price & Municipality

Price:

- Continuous numeric feature (float)
- Represents the sale or rental price of the apartment in BAM
- Serves as the target variable for our regression models

Price & Municipality

Price:

- Continuous numeric feature (float)
- Represents the sale or rental price of the apartment in BAM
- Serves as the target variable for our regression models

Municipality:

- Nominal categorical feature
- Indicates where the apartment is located
- Values include:
  - Sarajevo - Novi Grad
  - Sarajevo - Centar
  - Ilidža, Hadžići, etc.
- Useful for capturing regional price differences

Condition, Ad Type, Property Type

Condition:

- Nominal categorical
- Describes the apartment's condition (e.g., New Build, Renovated, Under Construction)
- Note: One noisy value is "Apartman" (misclassified entry)

Condition, Ad Type, Property Type

Condition:

- Nominal categorical
- Describes the apartment's condition (e.g., New Build, Renovated, Under Construction)
- Note: One noisy value is "Apartman" (misclassified entry)

Ad Type:

- Binary categorical
- Indicates whether the listing is for Sale or Rent
- Crucial for us, as we aim to build two separate models

Condition, Ad Type, Property Type

Condition:

- Nominal categorical
- Describes the apartment's condition (e.g., New Build, Renovated, Under Construction)
- Note: One noisy value is "Apartman" (misclassified entry)

Ad Type:

- Binary categorical
- Indicates whether the listing is for Sale or Rent
- Crucial for us, as we aim to build two separate models

Property Type:

- Binary categorical: Apartment vs. Flat
- Values are: *"Apartman"* or *"Stan"*

Rooms, Area, Furnishing Level

Number of Rooms:

- Integer feature
- Unlike some countries that use decimal points for half-rooms, Bosnian listings use whole numbers only

Rooms, Area, Furnishing Level

Number of Rooms:

- Integer feature
- Unlike some countries that use decimal points for half-rooms, Bosnian listings use whole numbers only

Area (m^2):

- Numeric continuous feature
- Indicates the size of the apartment in square meters

Rooms, Area, Furnishing Level

Number of Rooms:

- Integer feature
- Unlike some countries that use decimal points for half-rooms, Bosnian listings use whole numbers only

Area (m^2):

- Numeric continuous feature
- Indicates the size of the apartment in square meters

Furnishing Level:

- Ordinal categorical feature
- Captures whether the apartment is:
  - Furnished
  - Partially furnished
  - Unfurnished

Floor Level & Heating Type

Floor Level:

- Mostly integer values (e.g., 0, -2, 3, 7, 11)
- Some values are non-standard: *"20+"*
- Could be treated as:
  - Ordinal (if treated as relative height)
  - Categorical (if "20+" cannot be normalized)

Floor Level & Heating Type

Floor Level:

- Mostly integer values (e.g., 0, -2, 3, 7, 11)
- Some values are non-standard: *"20+"*
- Could be treated as:
  - Ordinal (if treated as relative height)
  - Categorical (if "20+" cannot be normalized)

Heating Type:

- Nominal categorical
- Examples:
  - City Central Heating
  - Electric Heating
  - Gas Heating
  - Wood
  - Other
- May affect price and energy efficiency rating

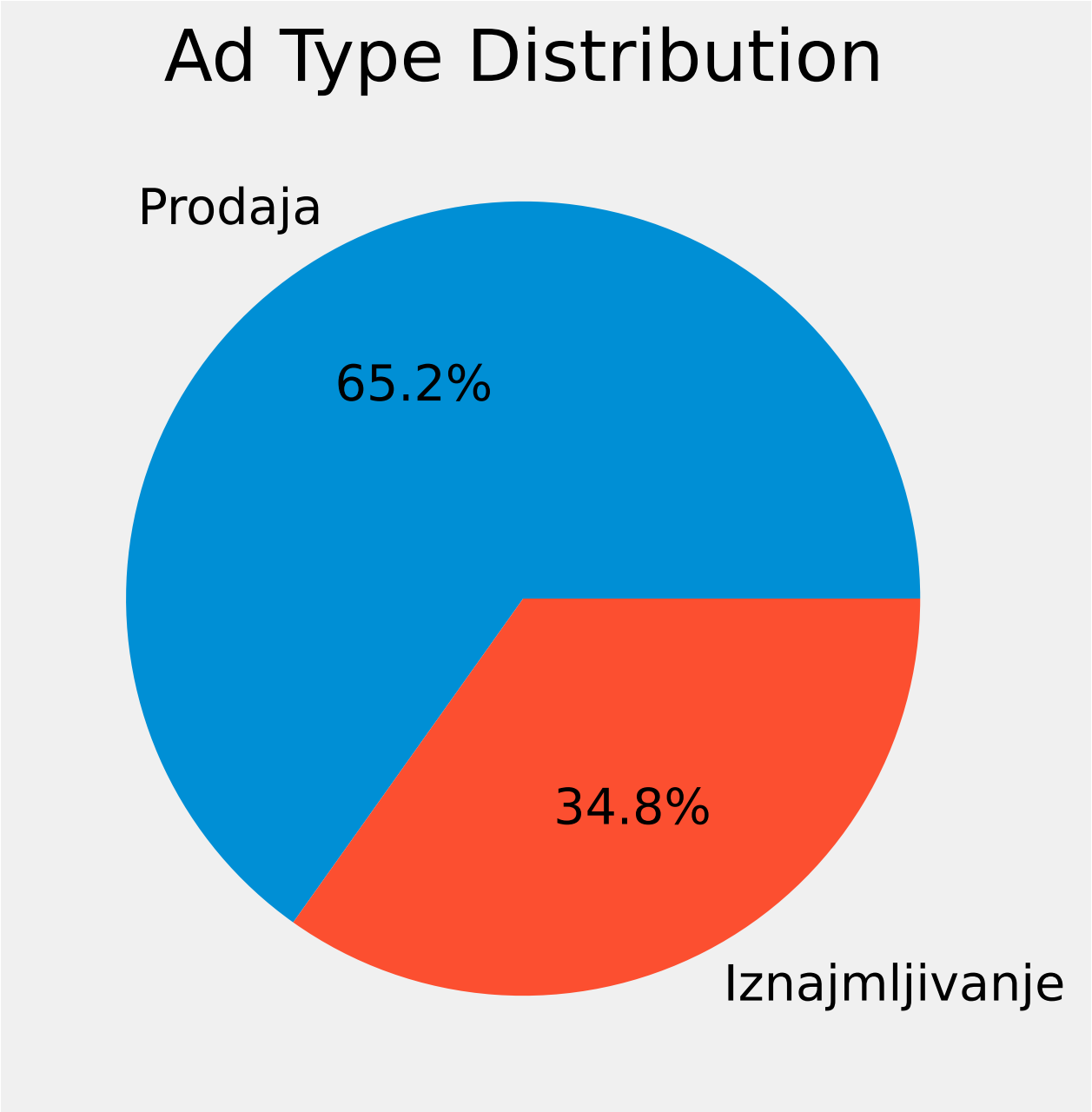
Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

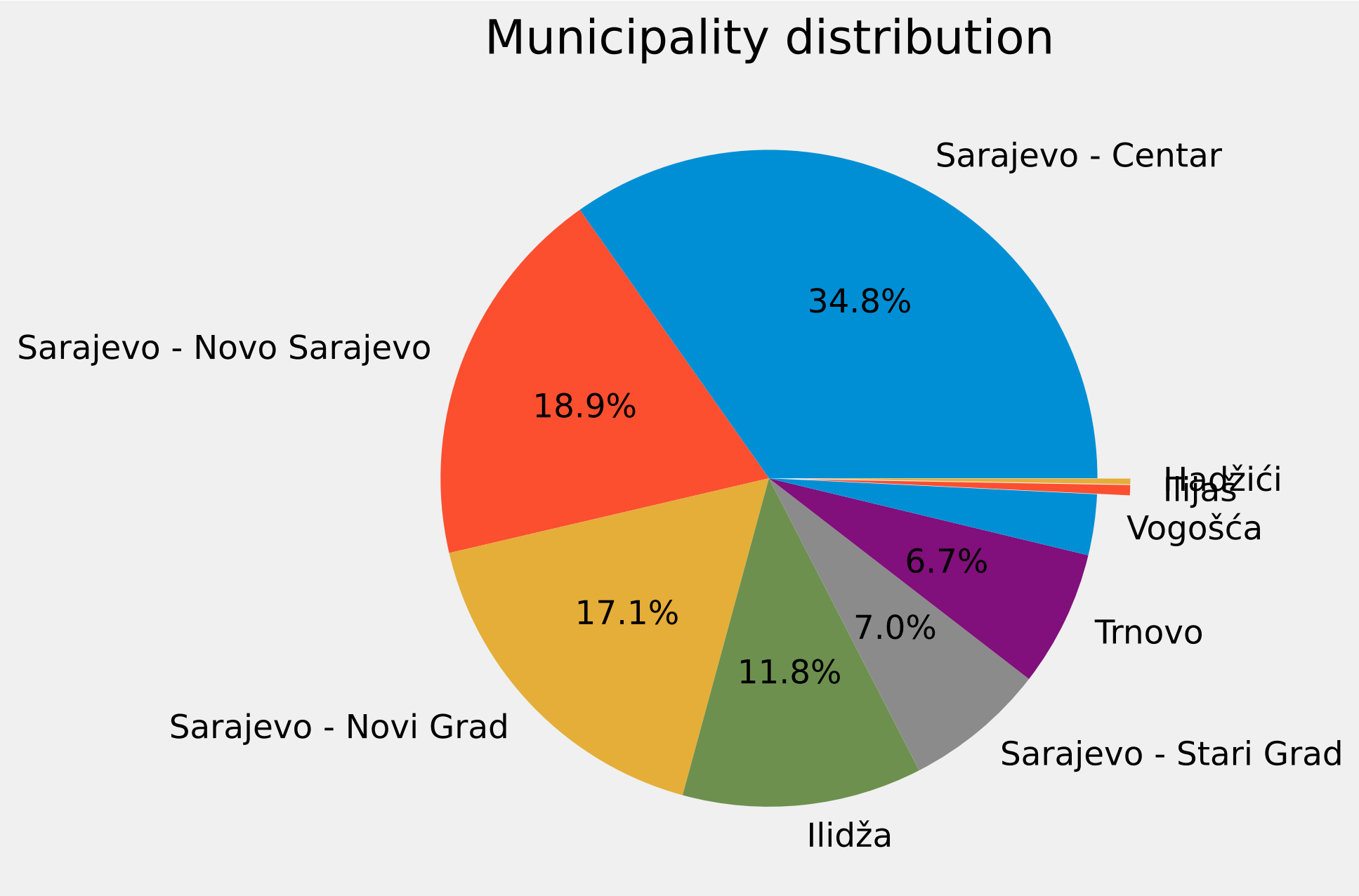
Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

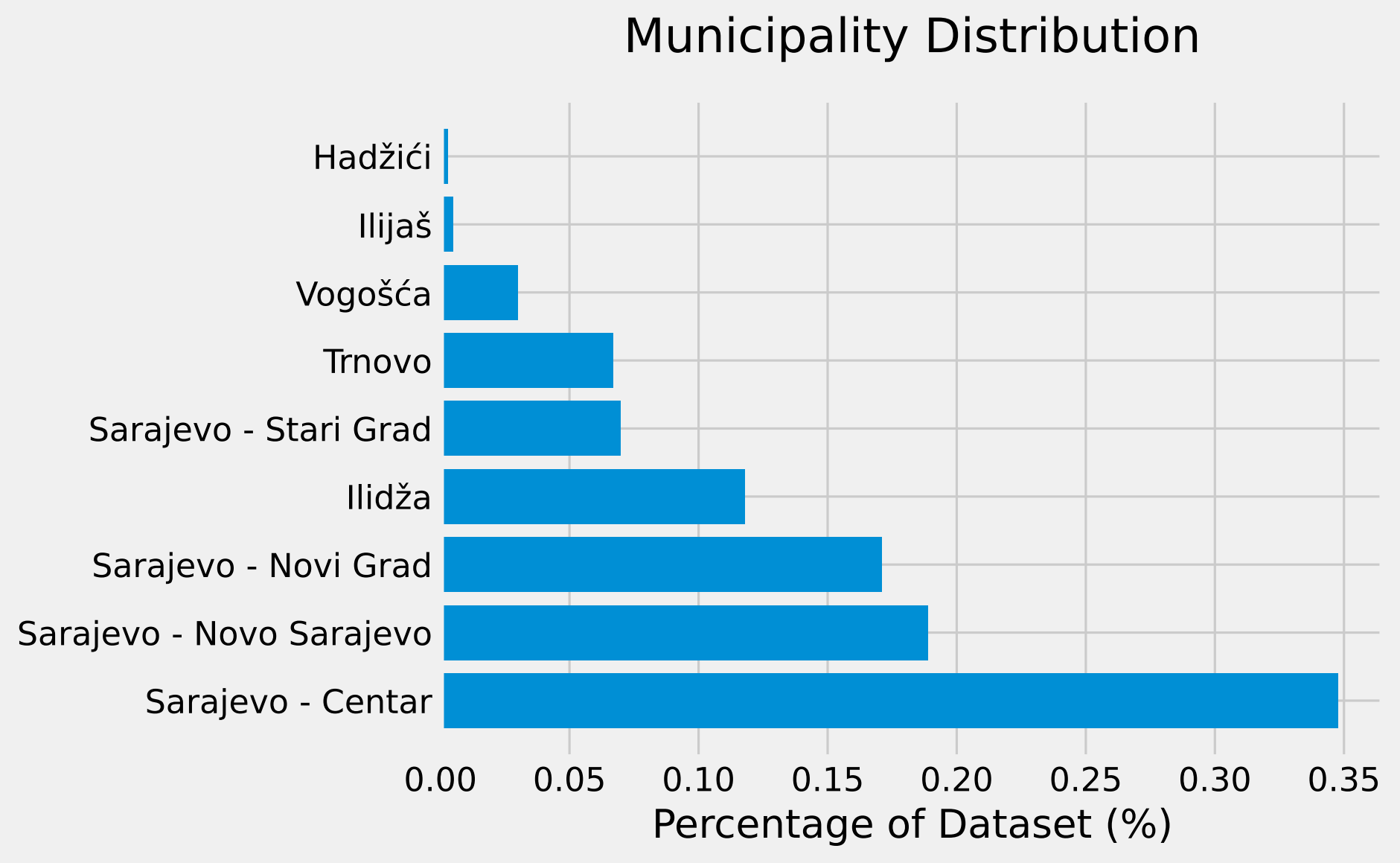
Ad Type Distribution



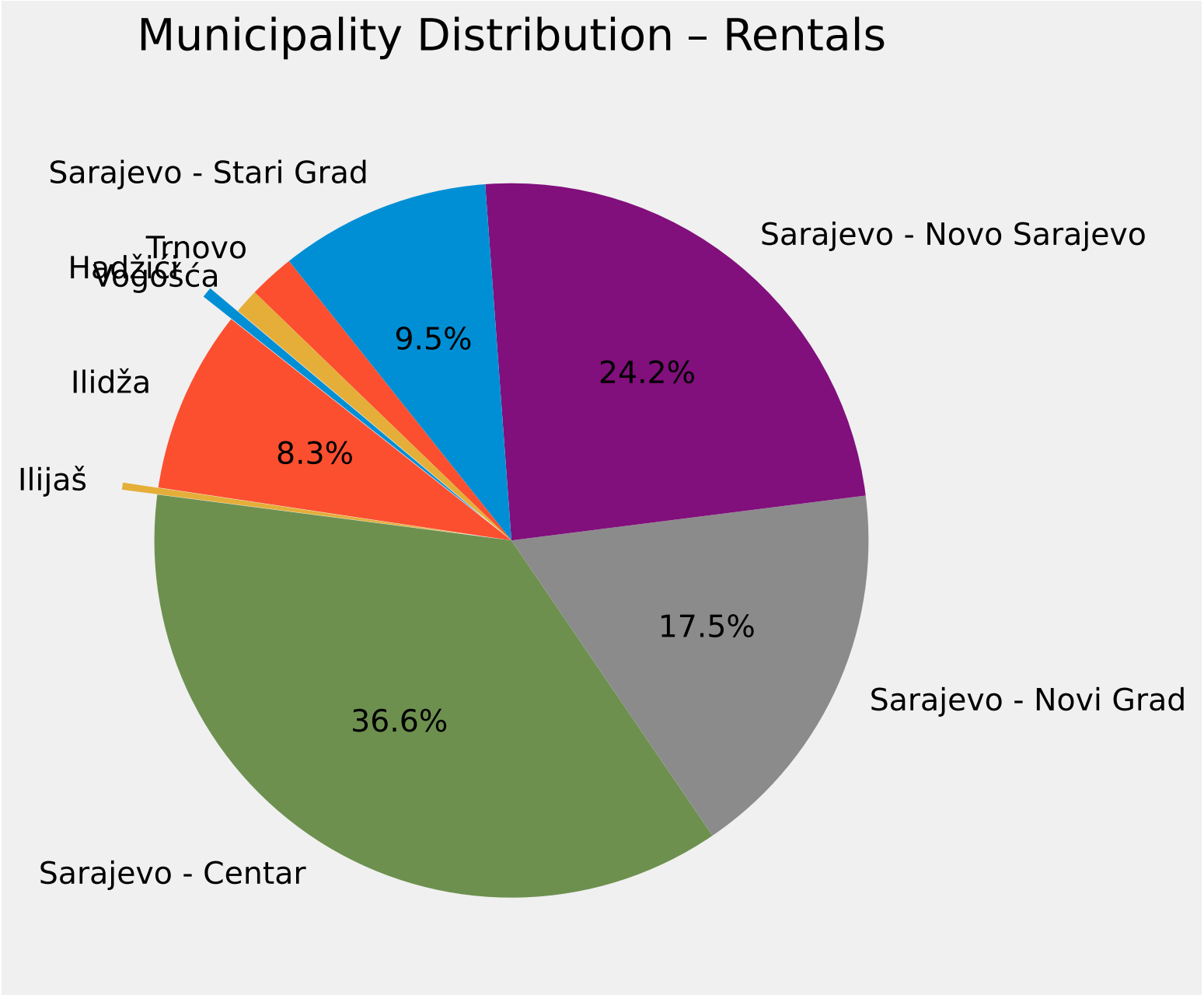
Municipality Distribution (Pie)



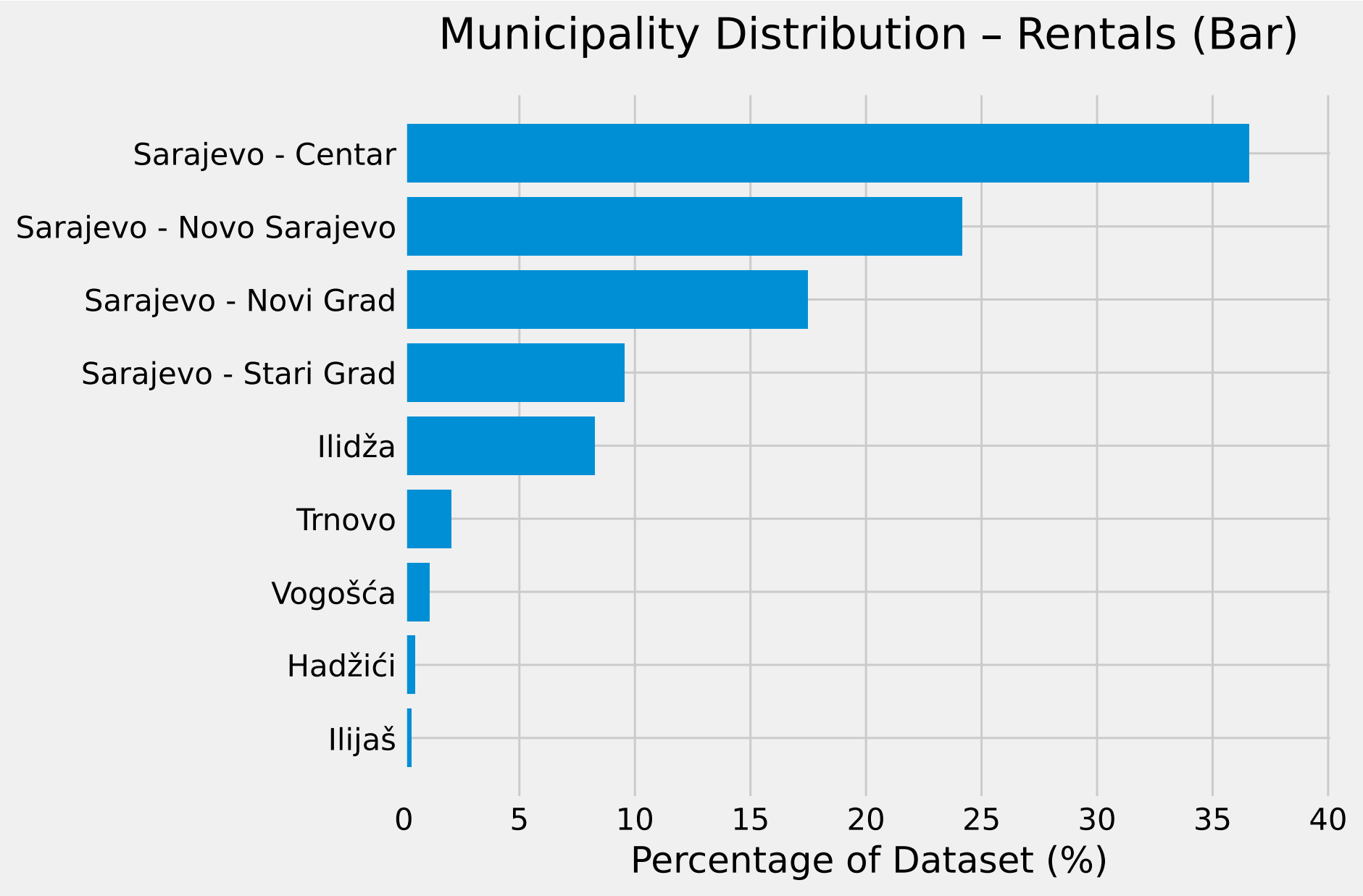
Municipality Distribution (Bar)



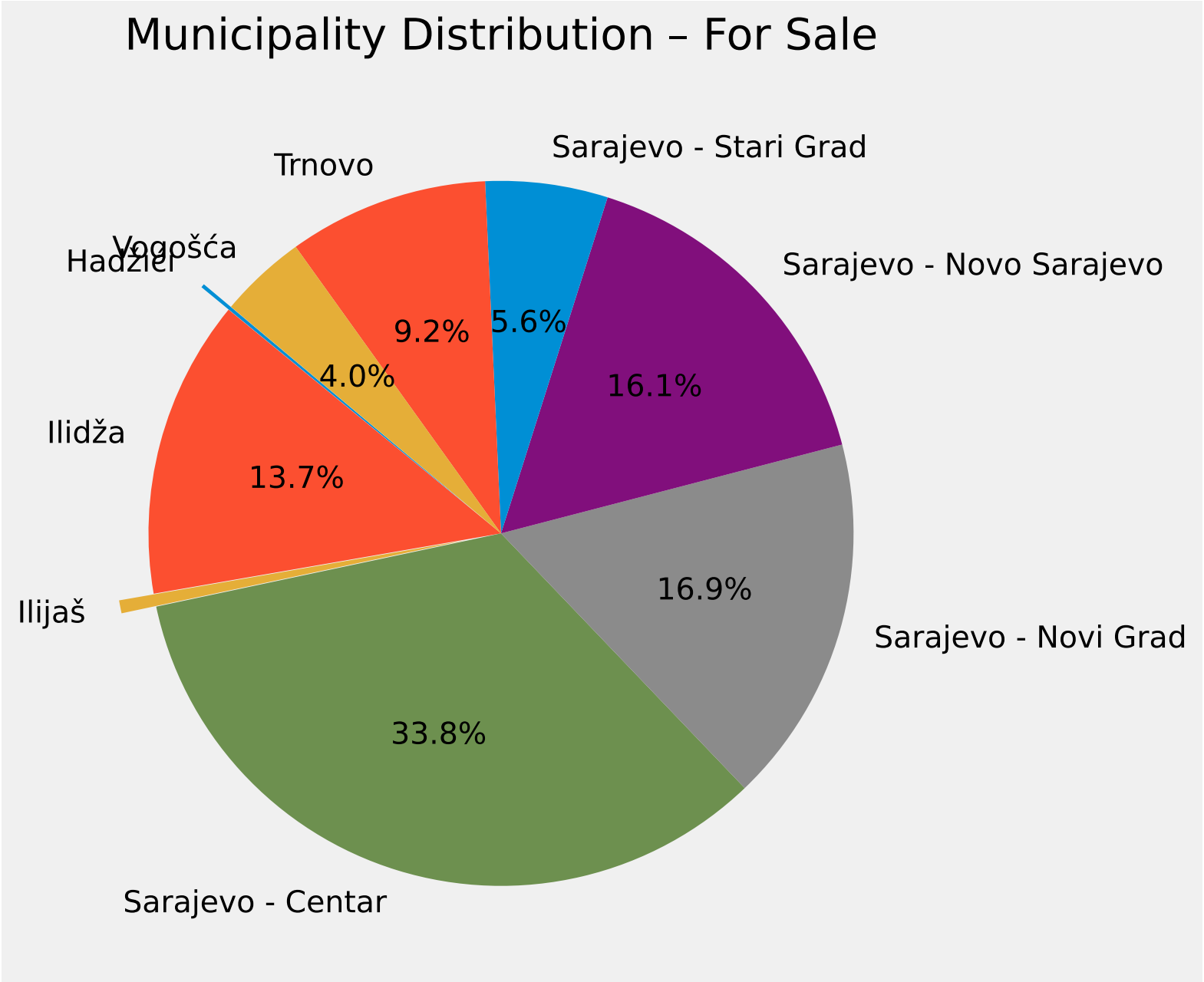
Municipality - Rentals (Pie)



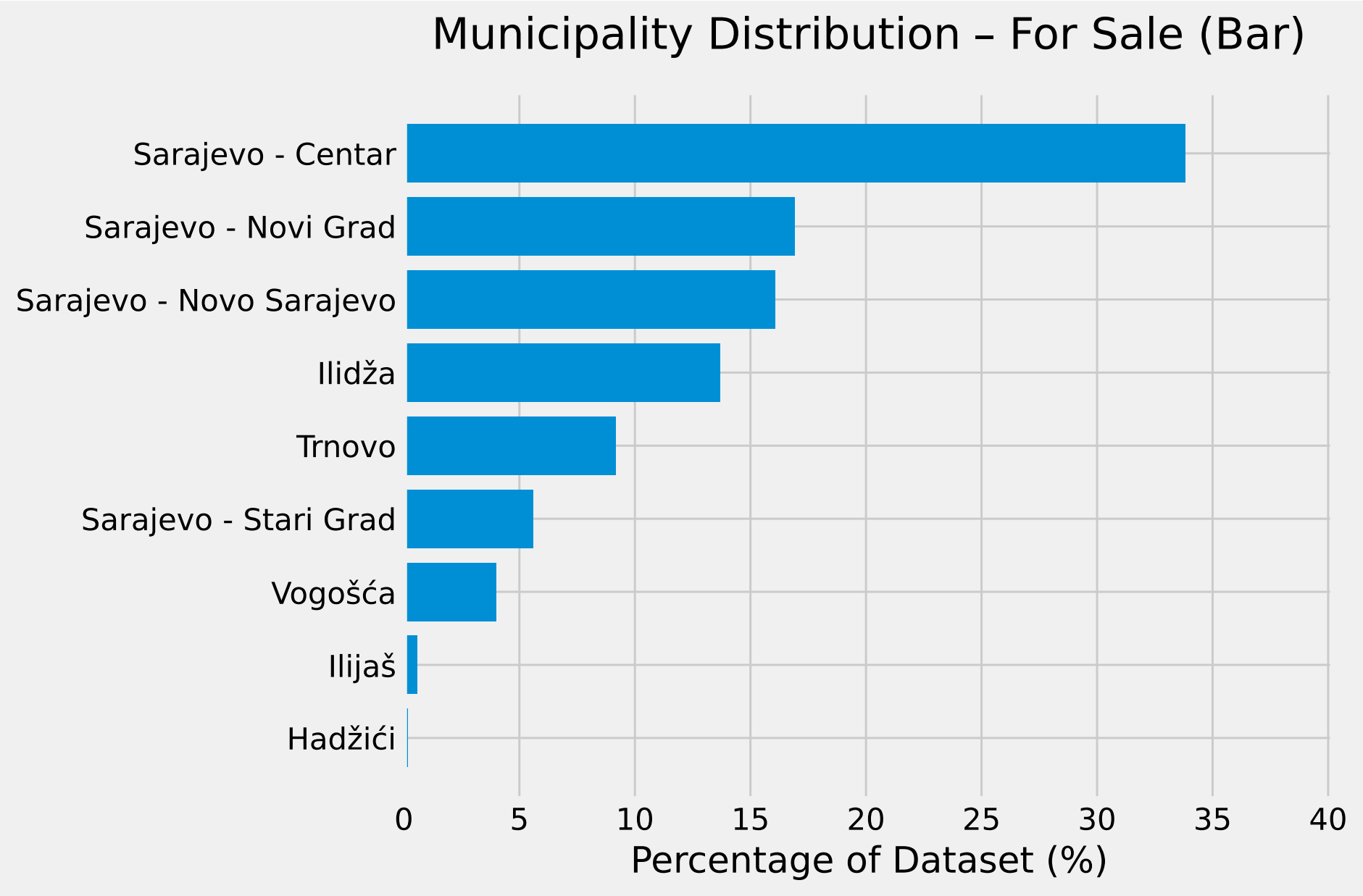
Municipality - Rentals (Bar)



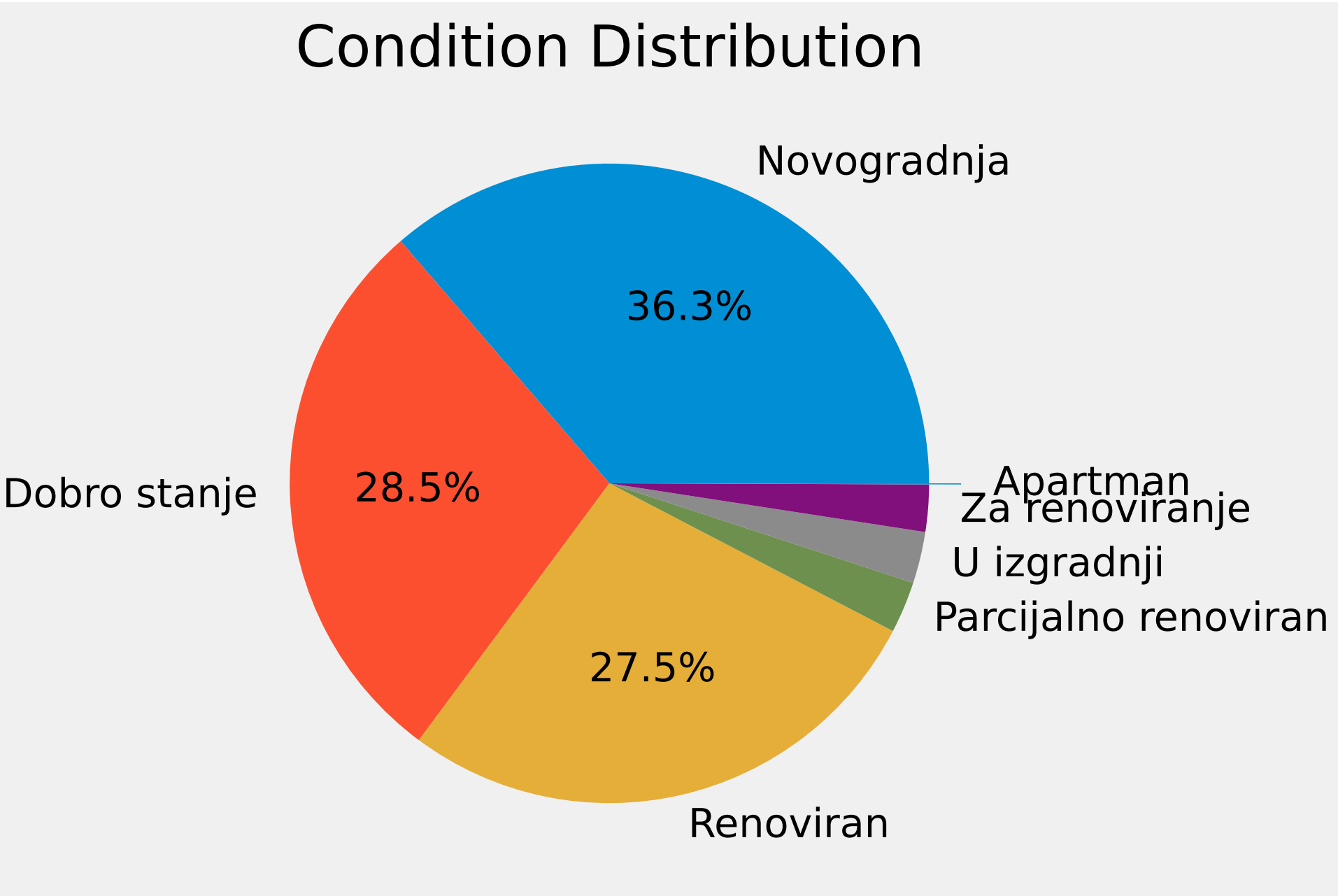
Municipality - For Sale (Pie)



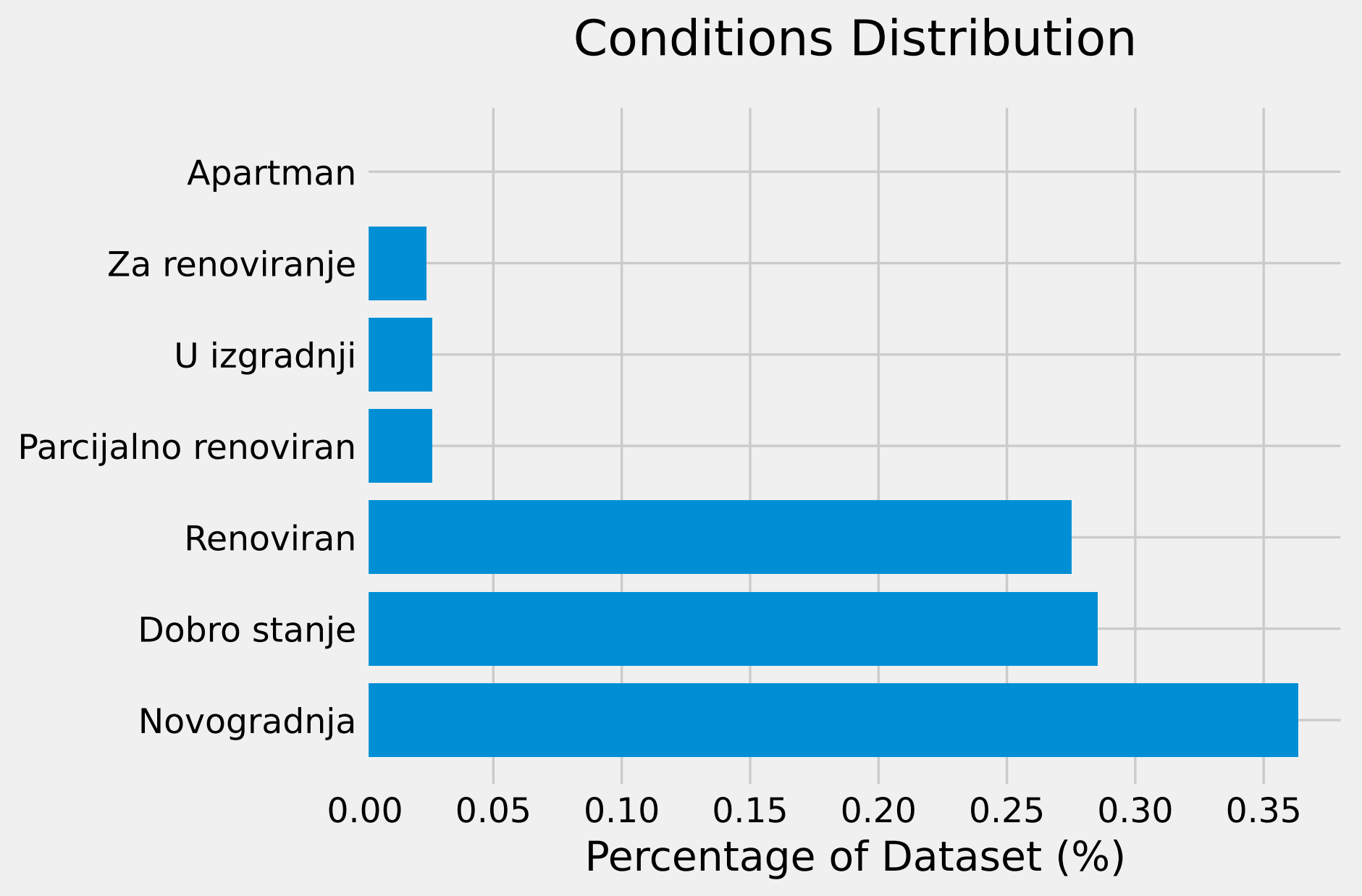
Municipality - For Sale (Bar)



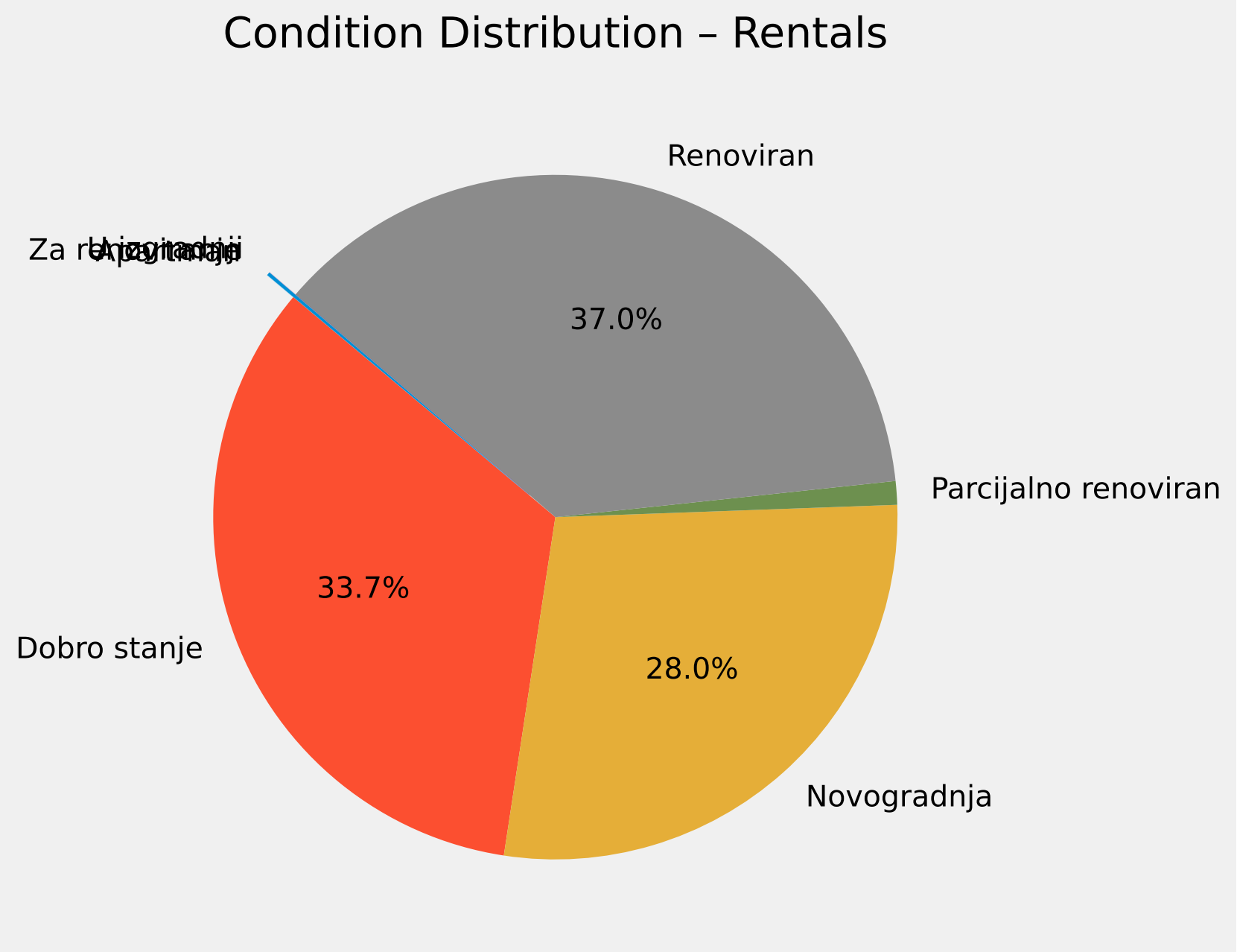
Conditions (Pie)



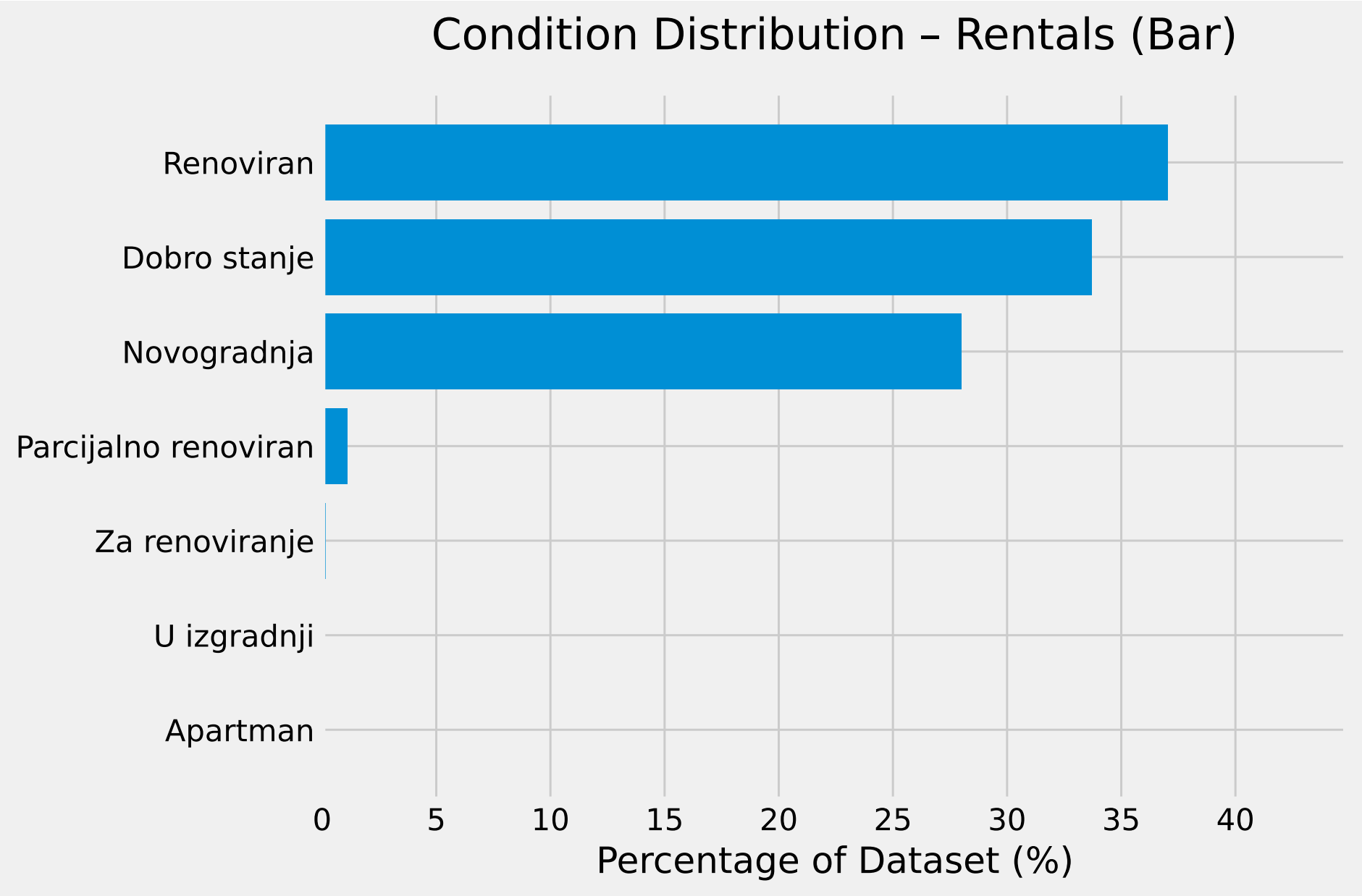
Conditions (Bar)



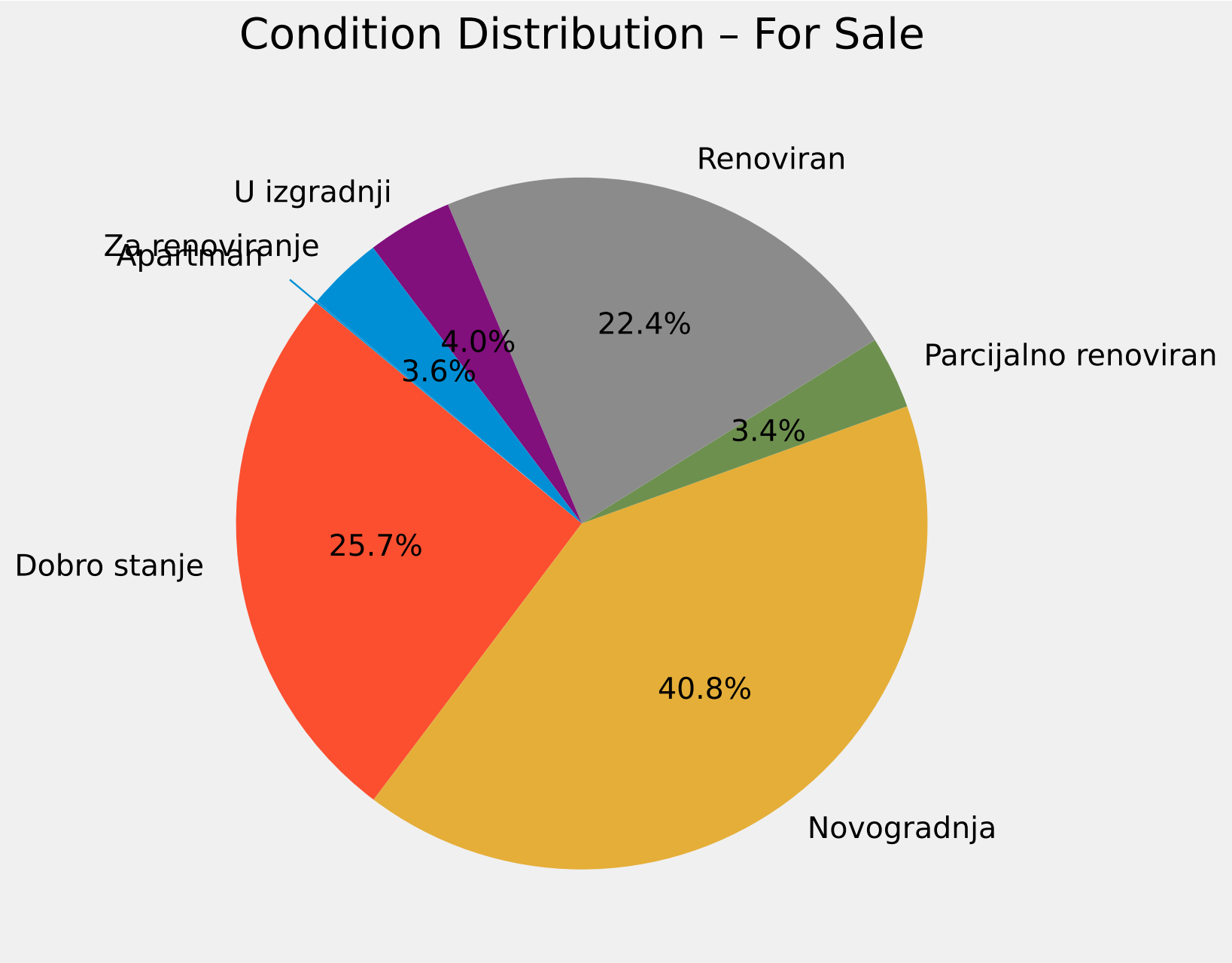
Conditions - Rentals (Pie)



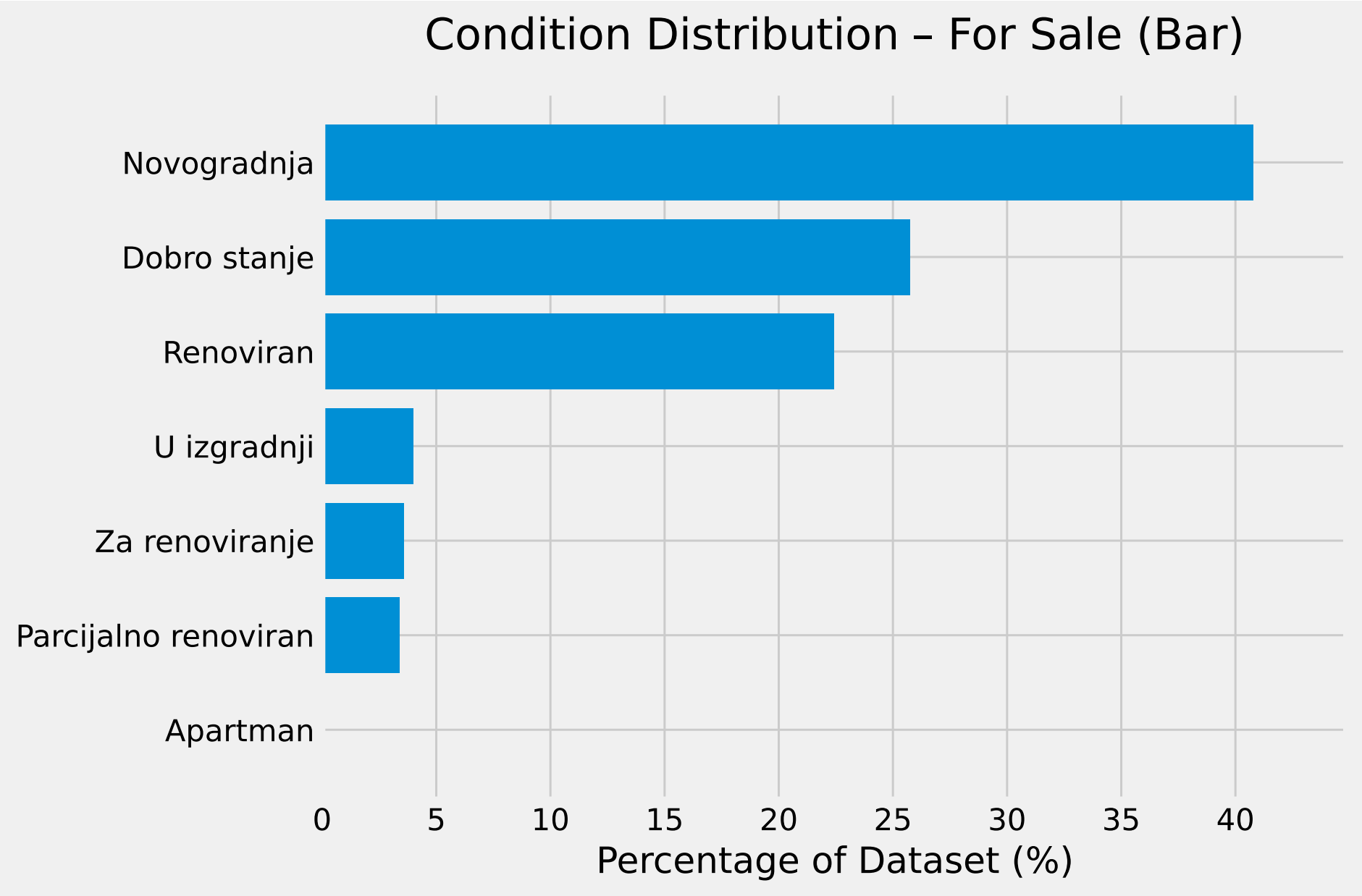
Conditions - Rentals (Bar)



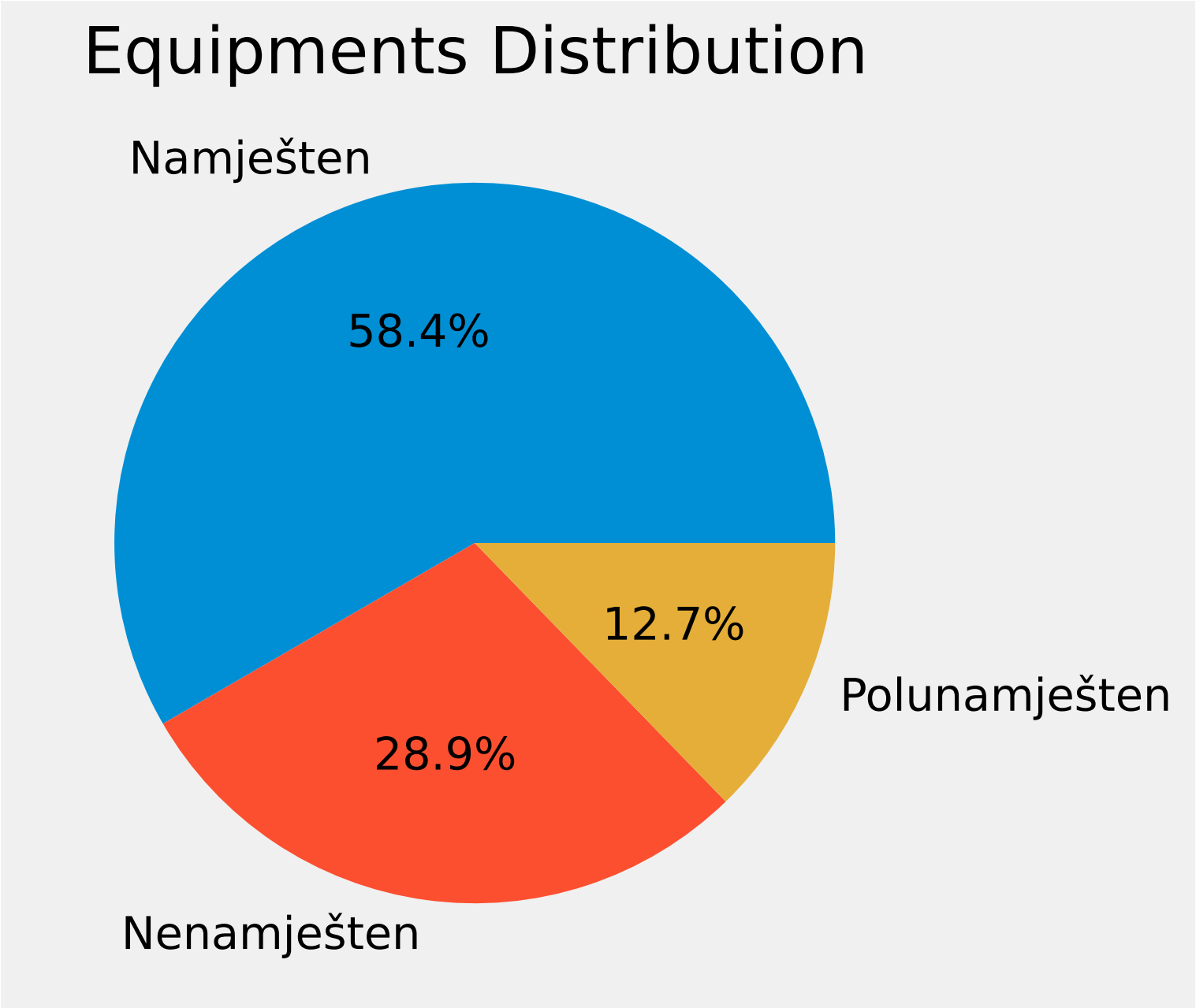
Conditions - For Sale (Pie)



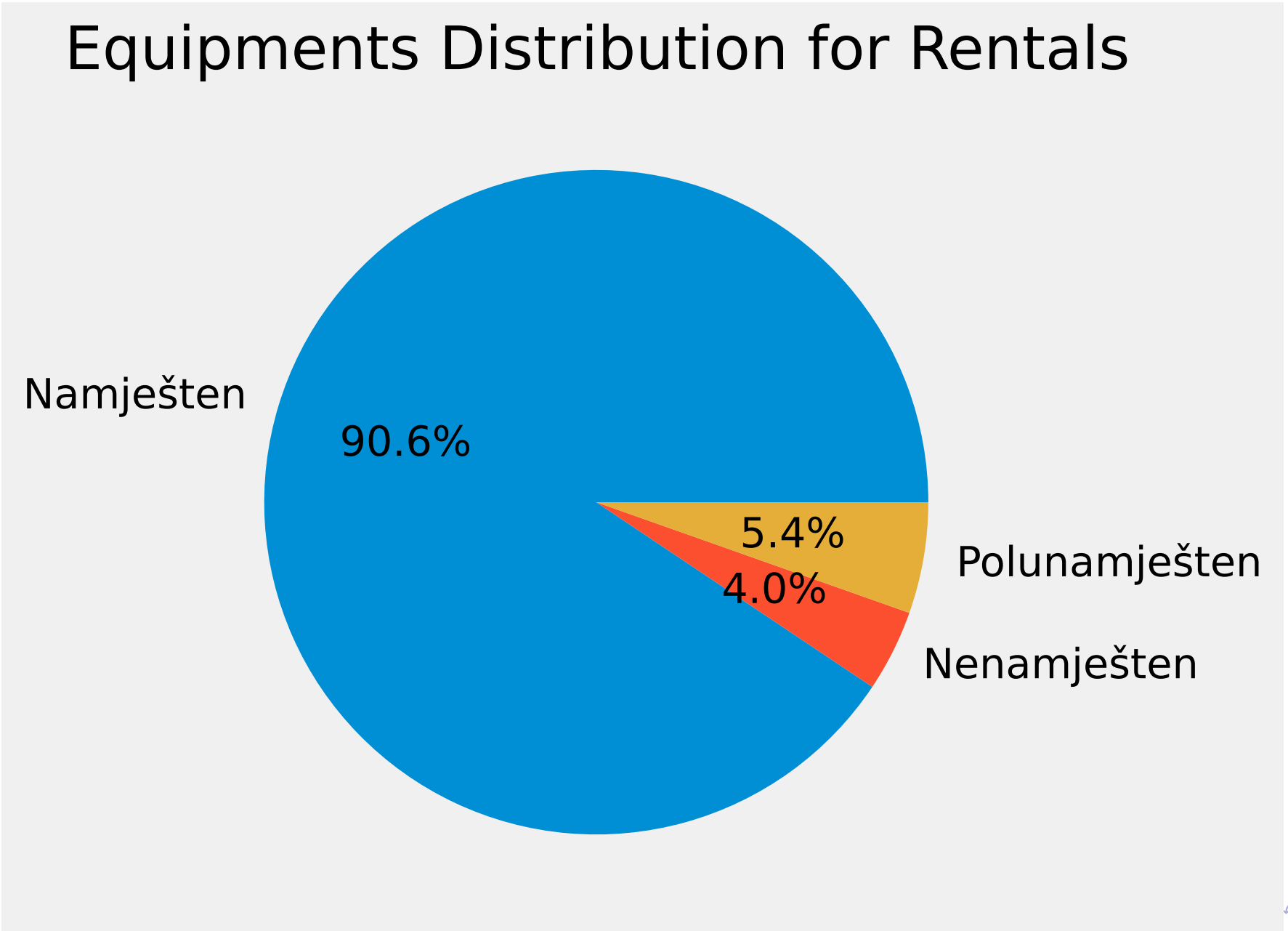
Conditions - For Sale (Bar)



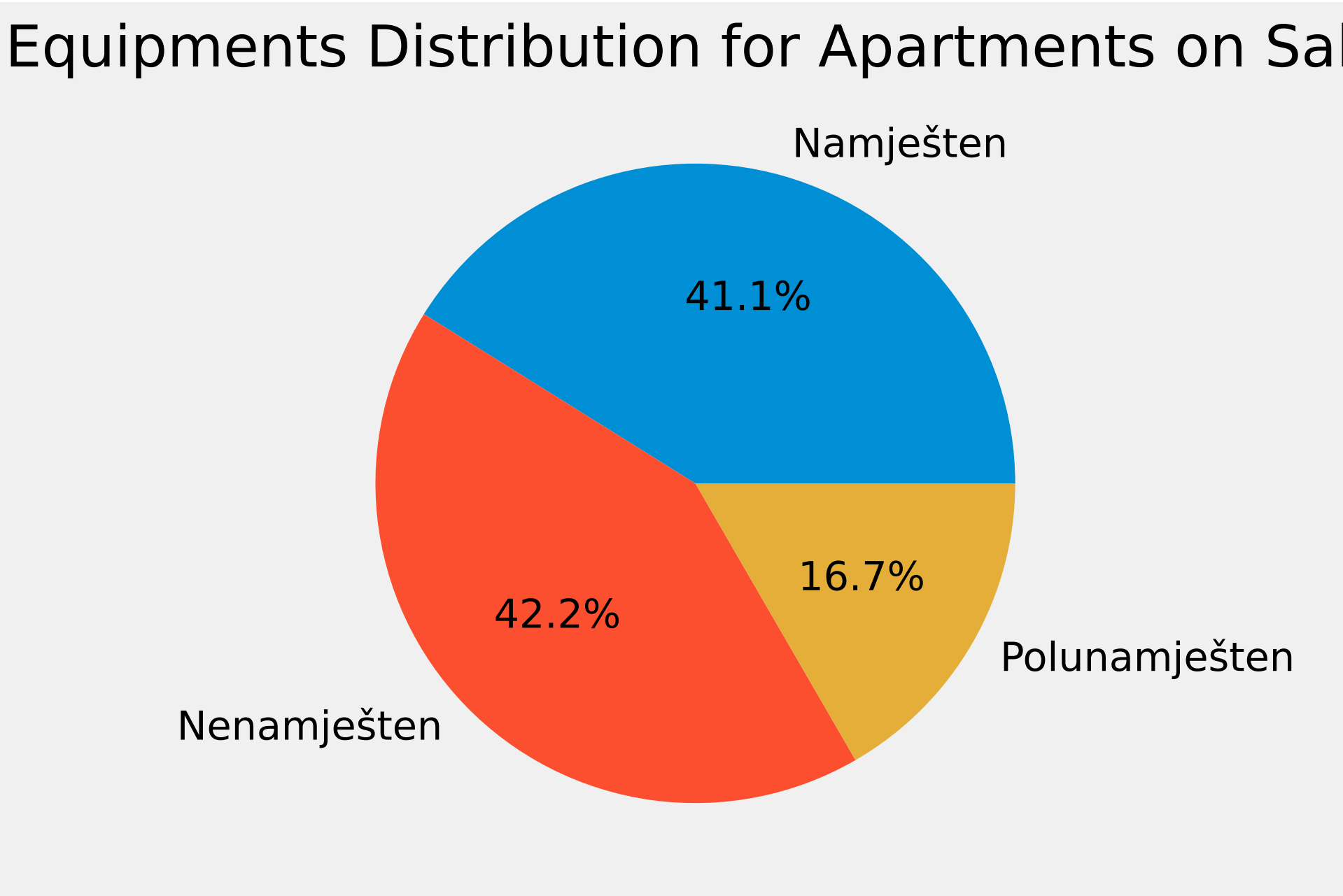
Equipment (All)



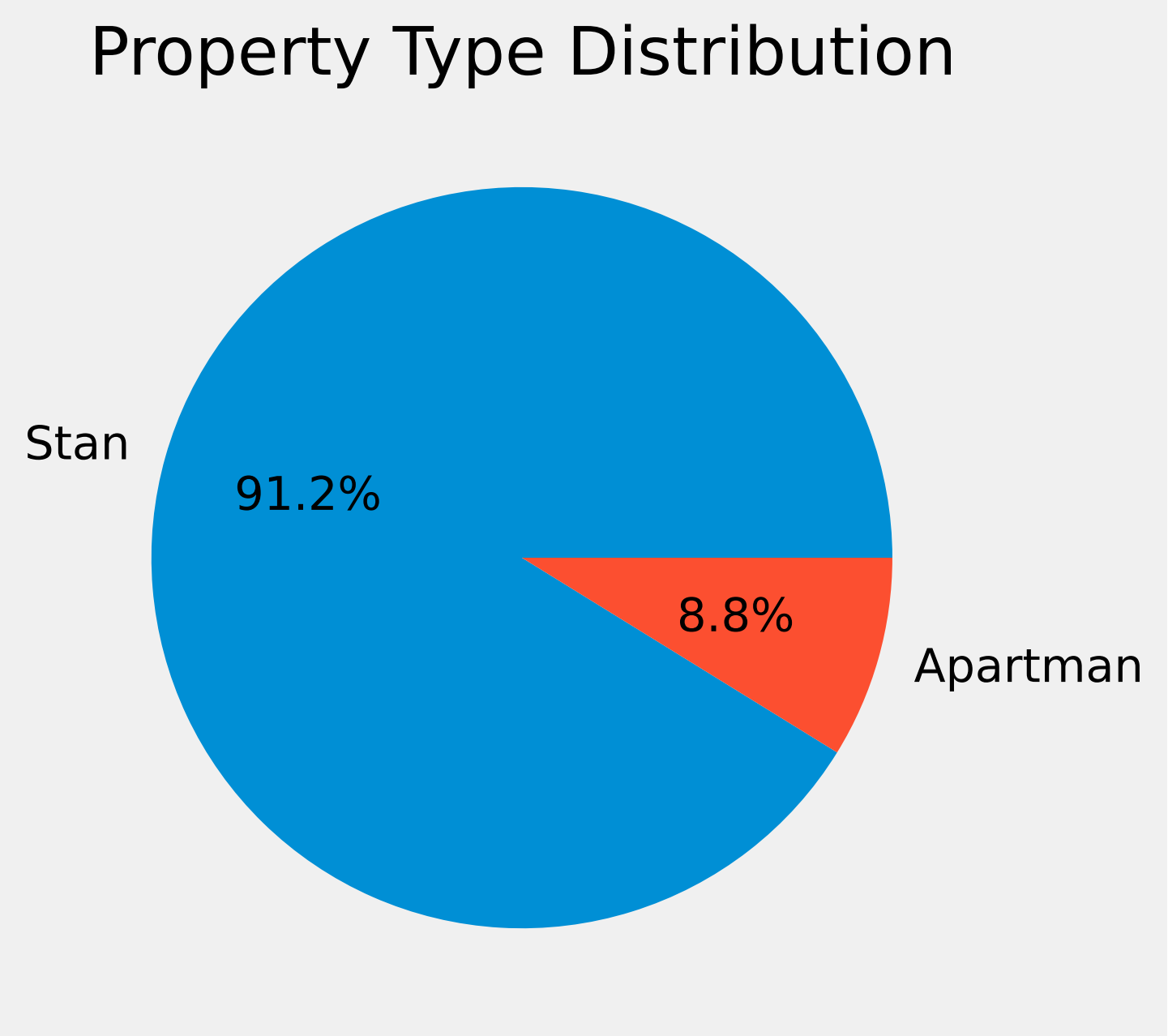
Equipment - Rentals



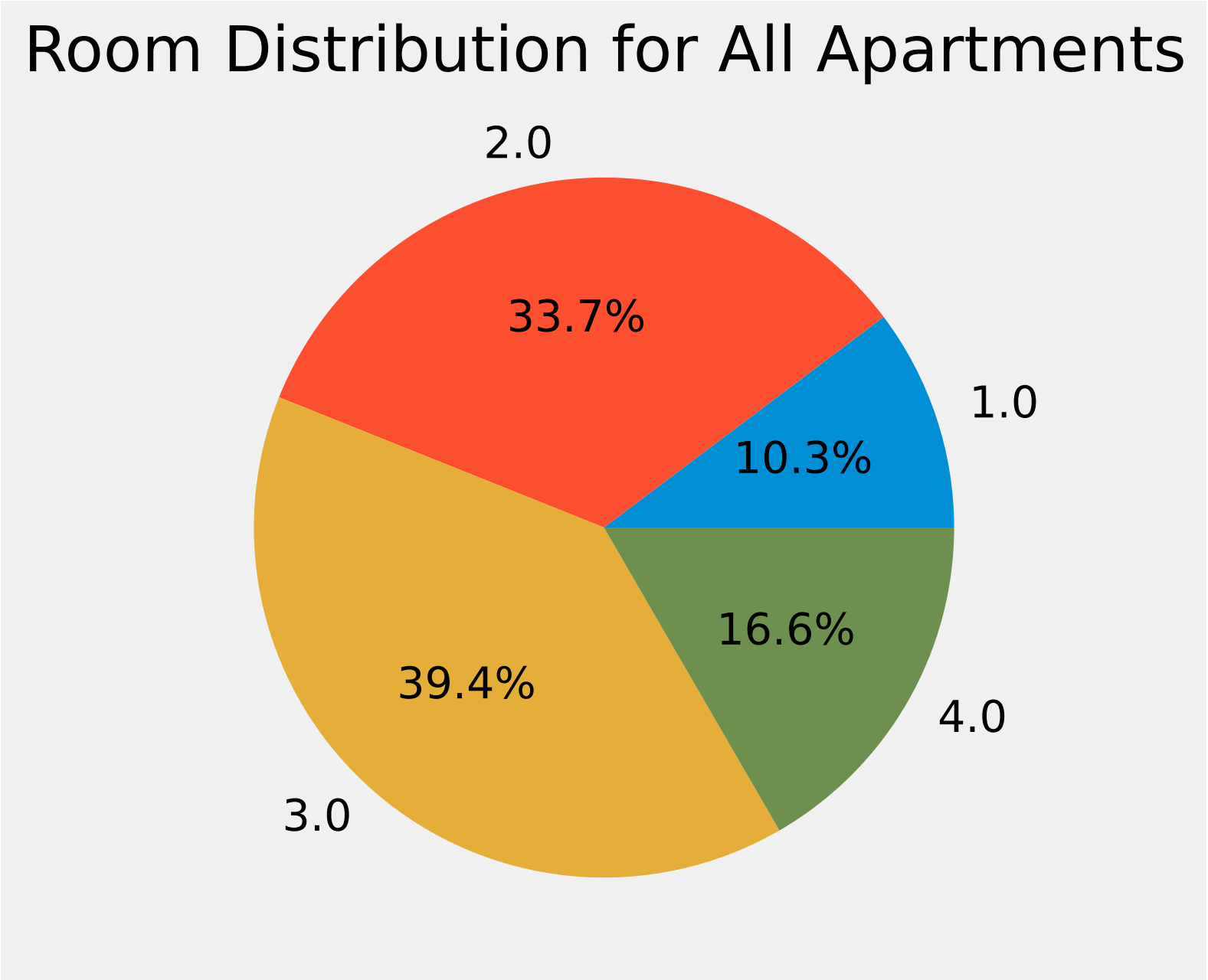
Equipment - For Sale



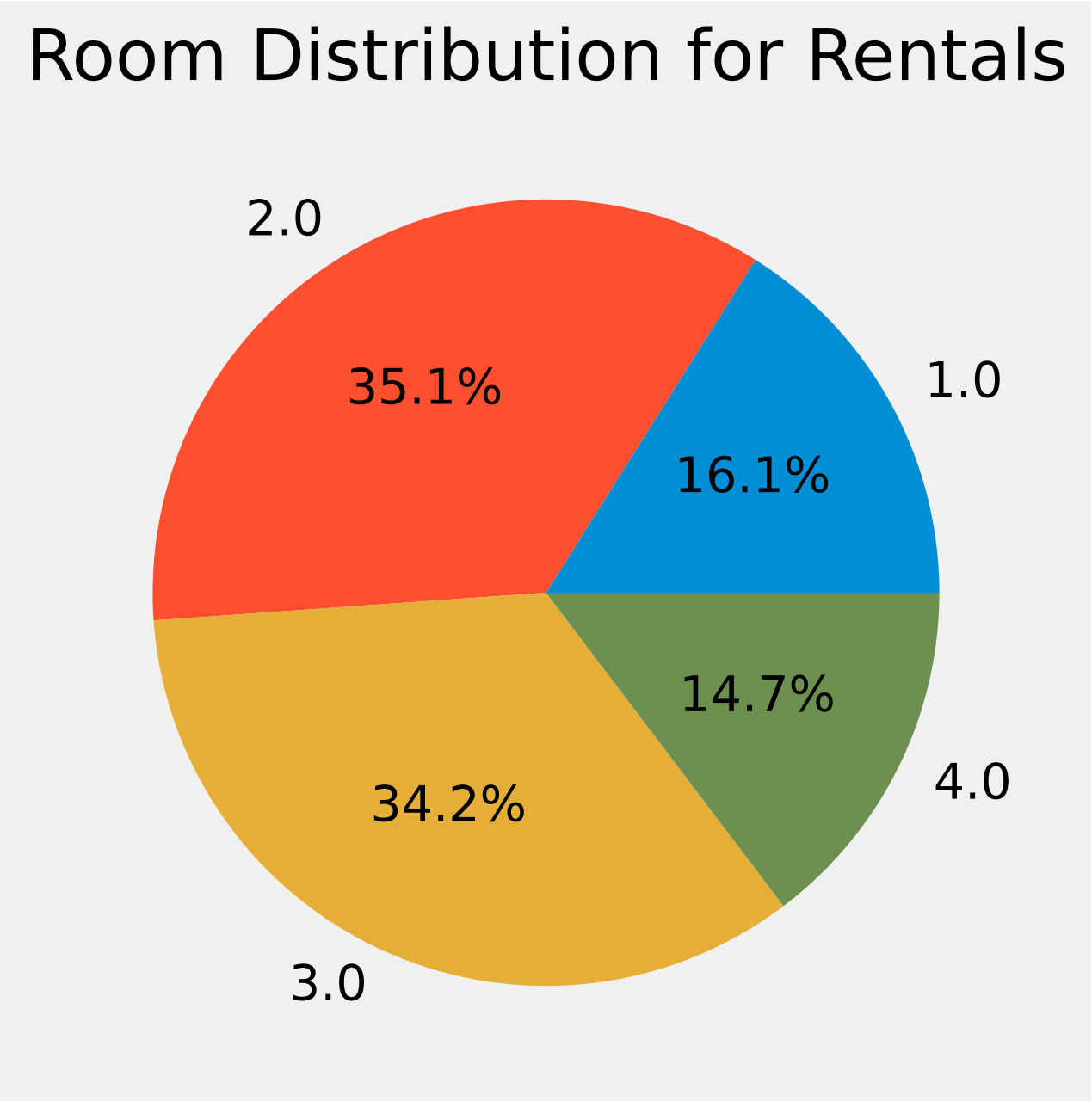
Property Type



Room Count - All

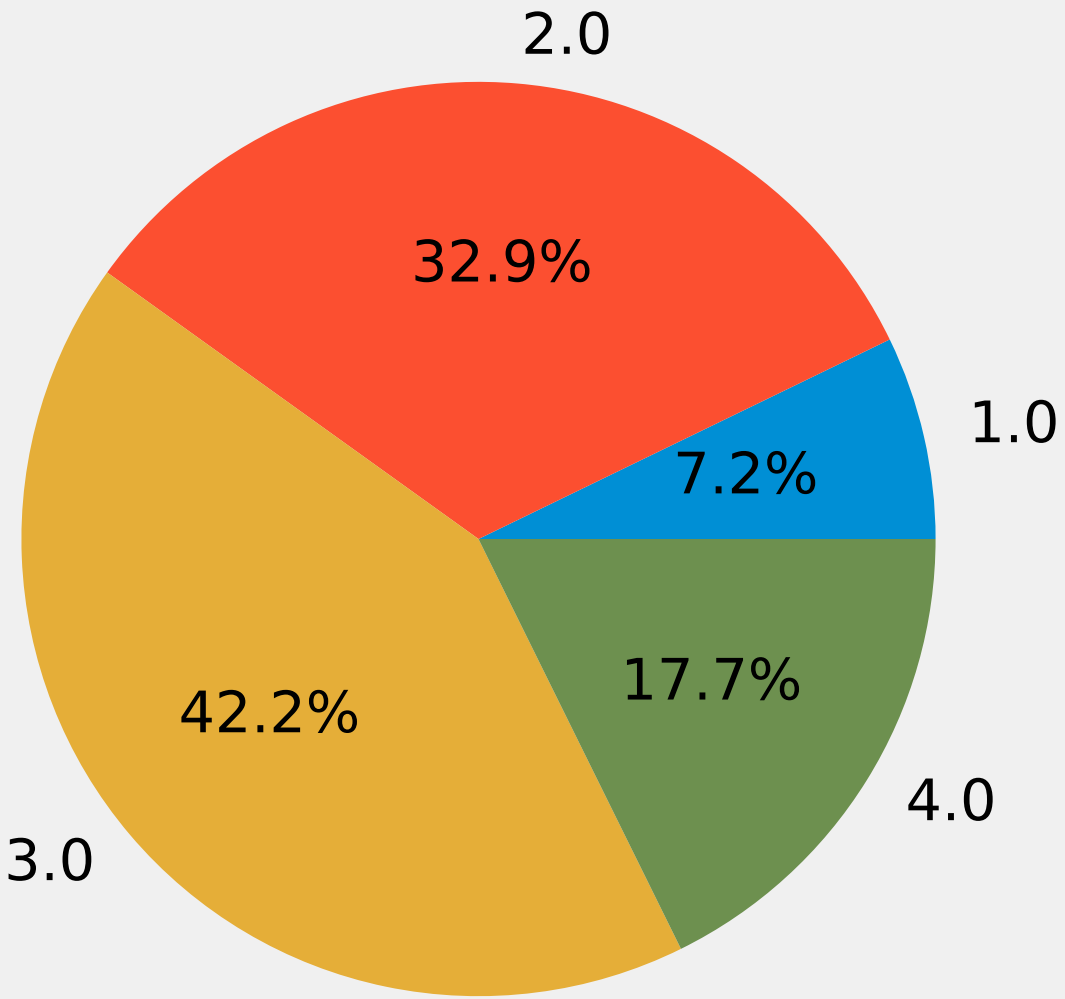


Room Count - Rentals



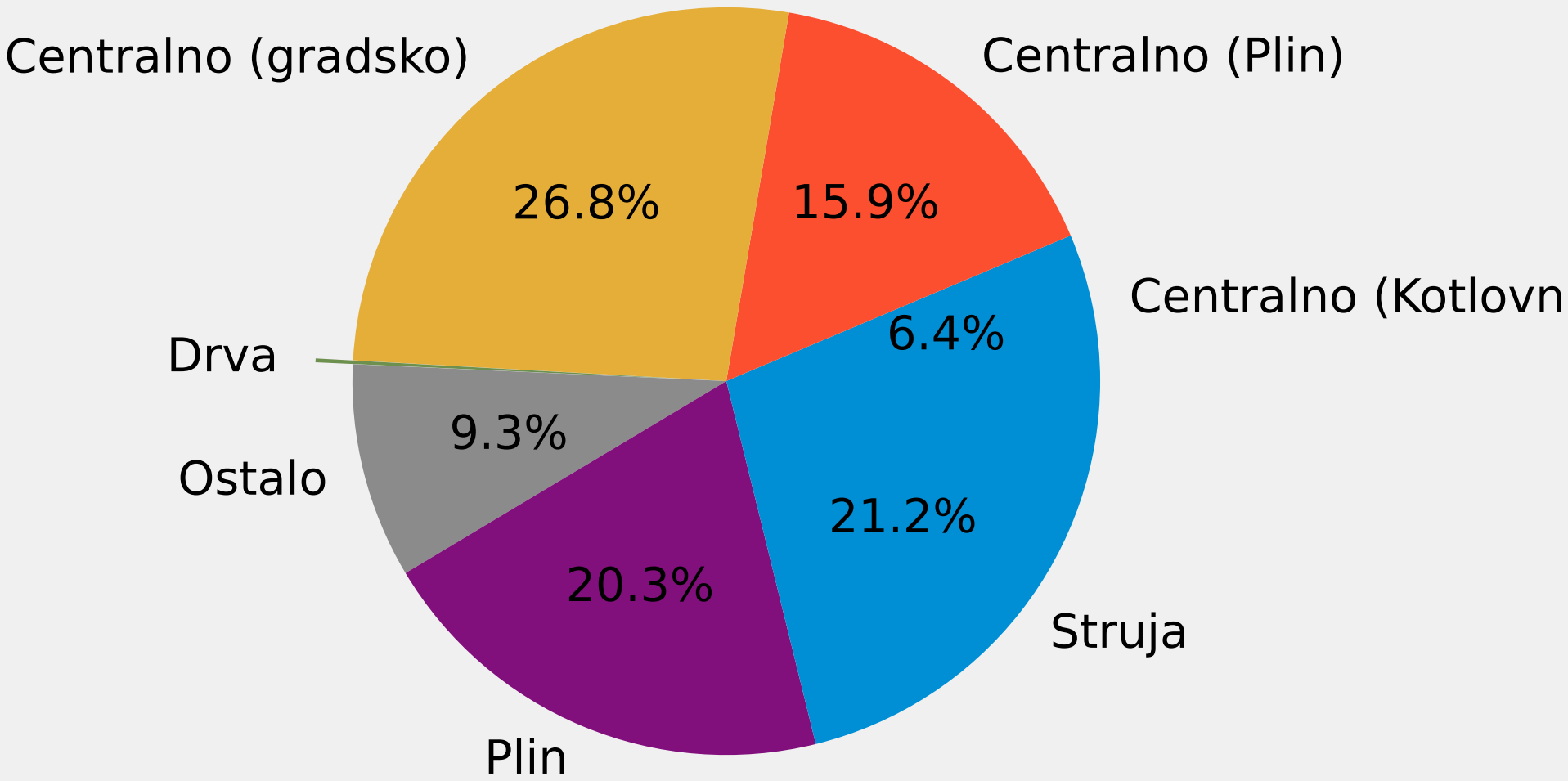
Room Count - For Sale

Room Distribution for Apartments on Sale

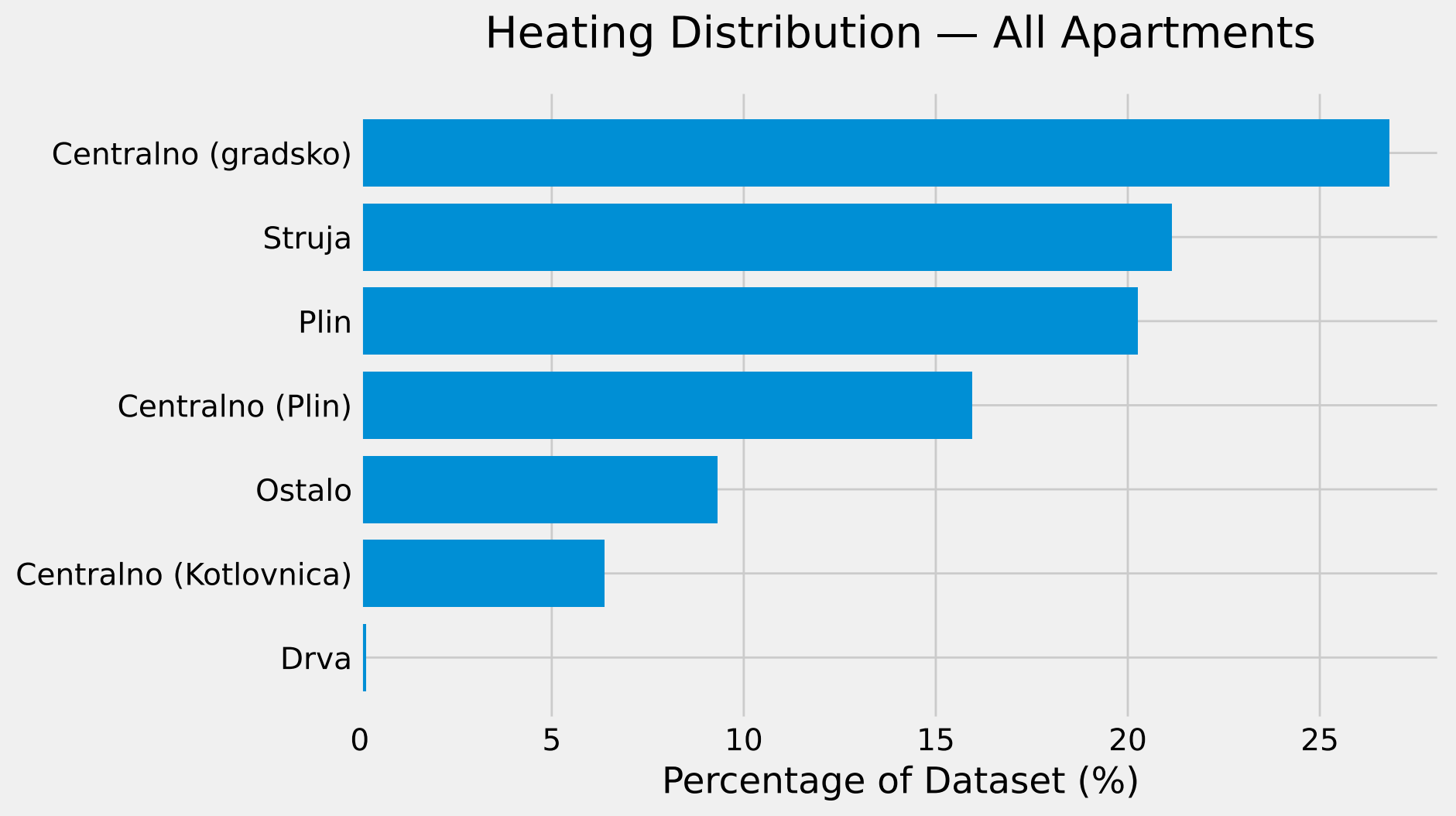


Heating - All (Pie)

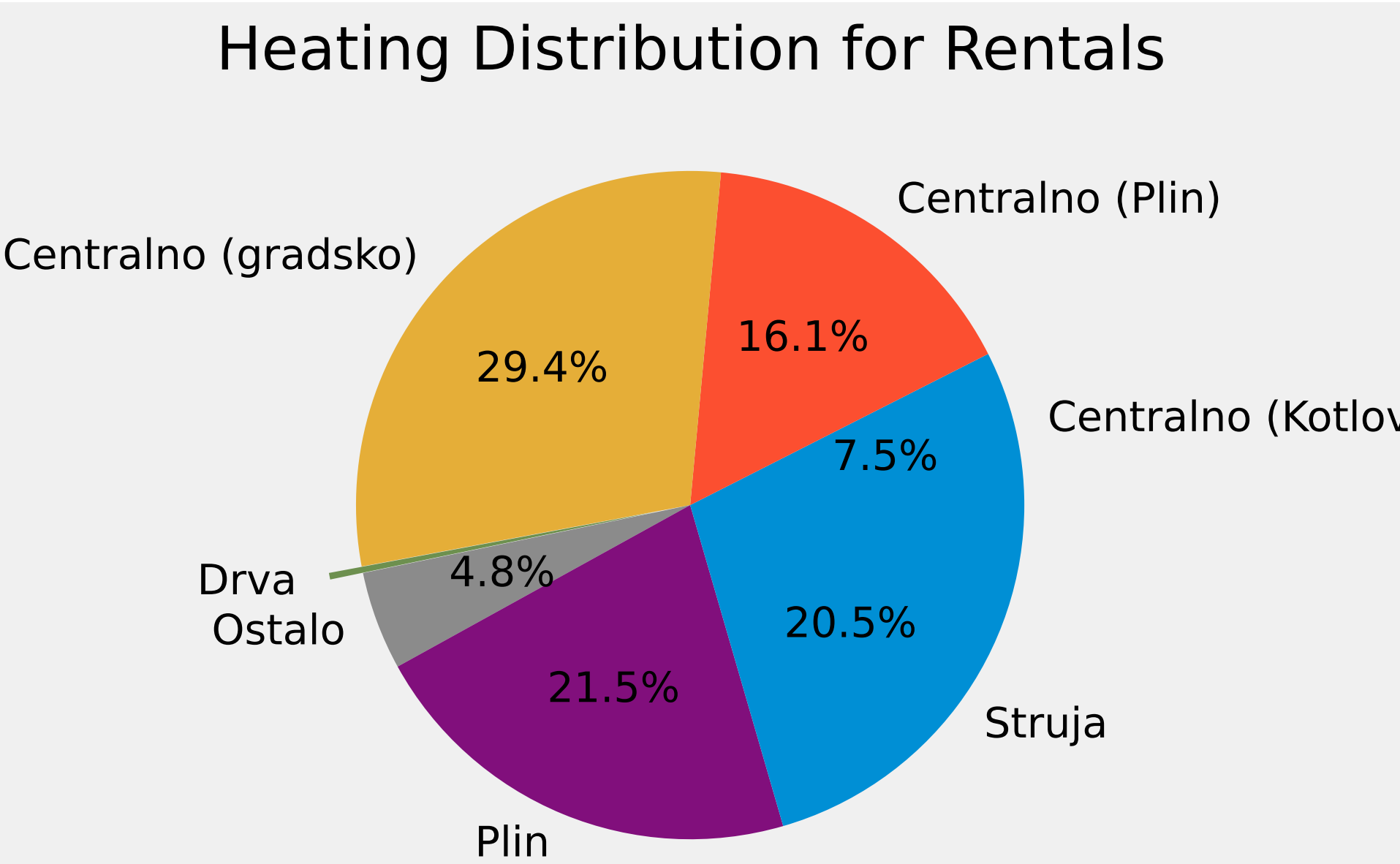
Heating Distribution for All Apartments



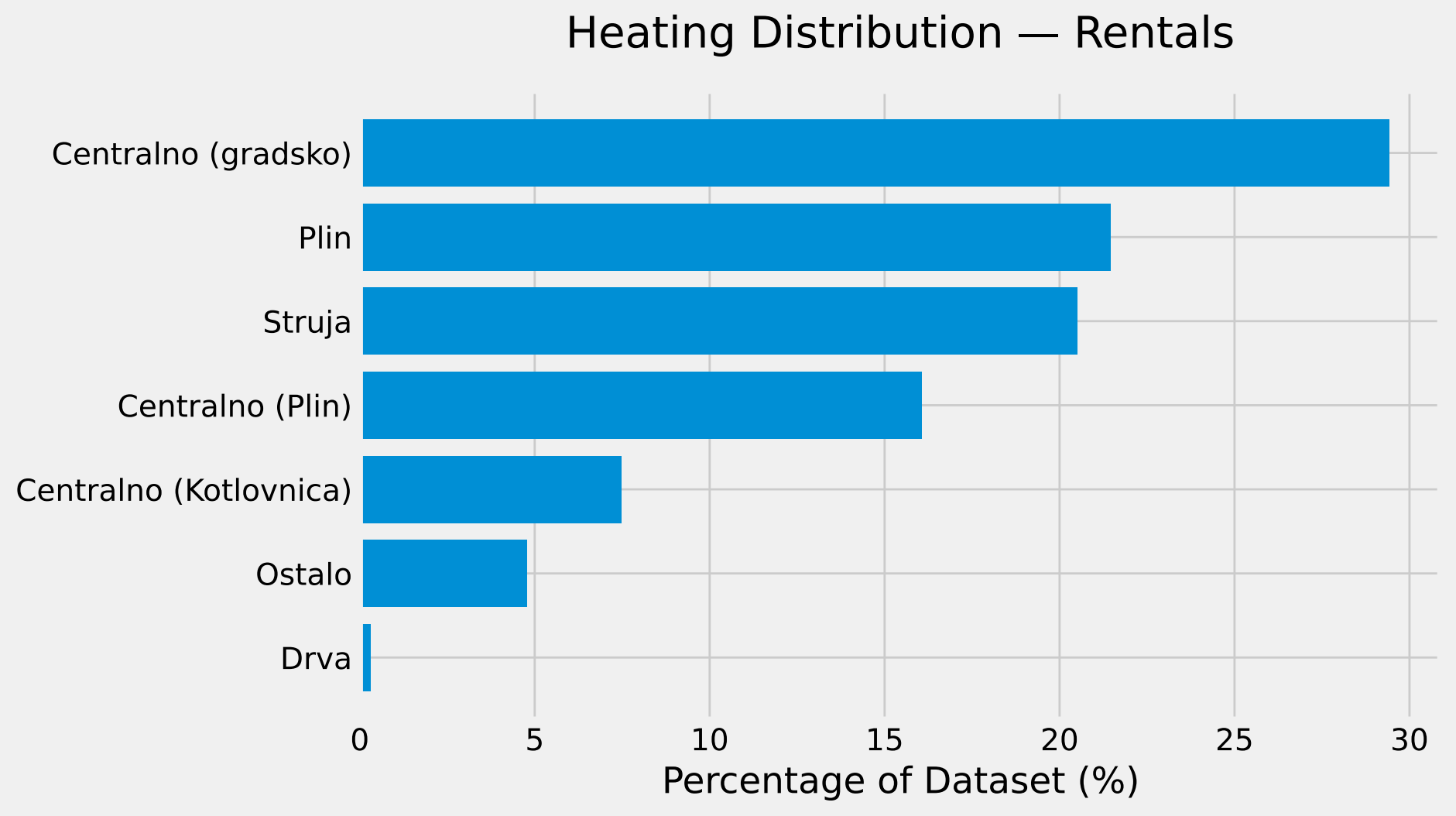
Heating - All (Bar)



Heating - Rentals (Pie)

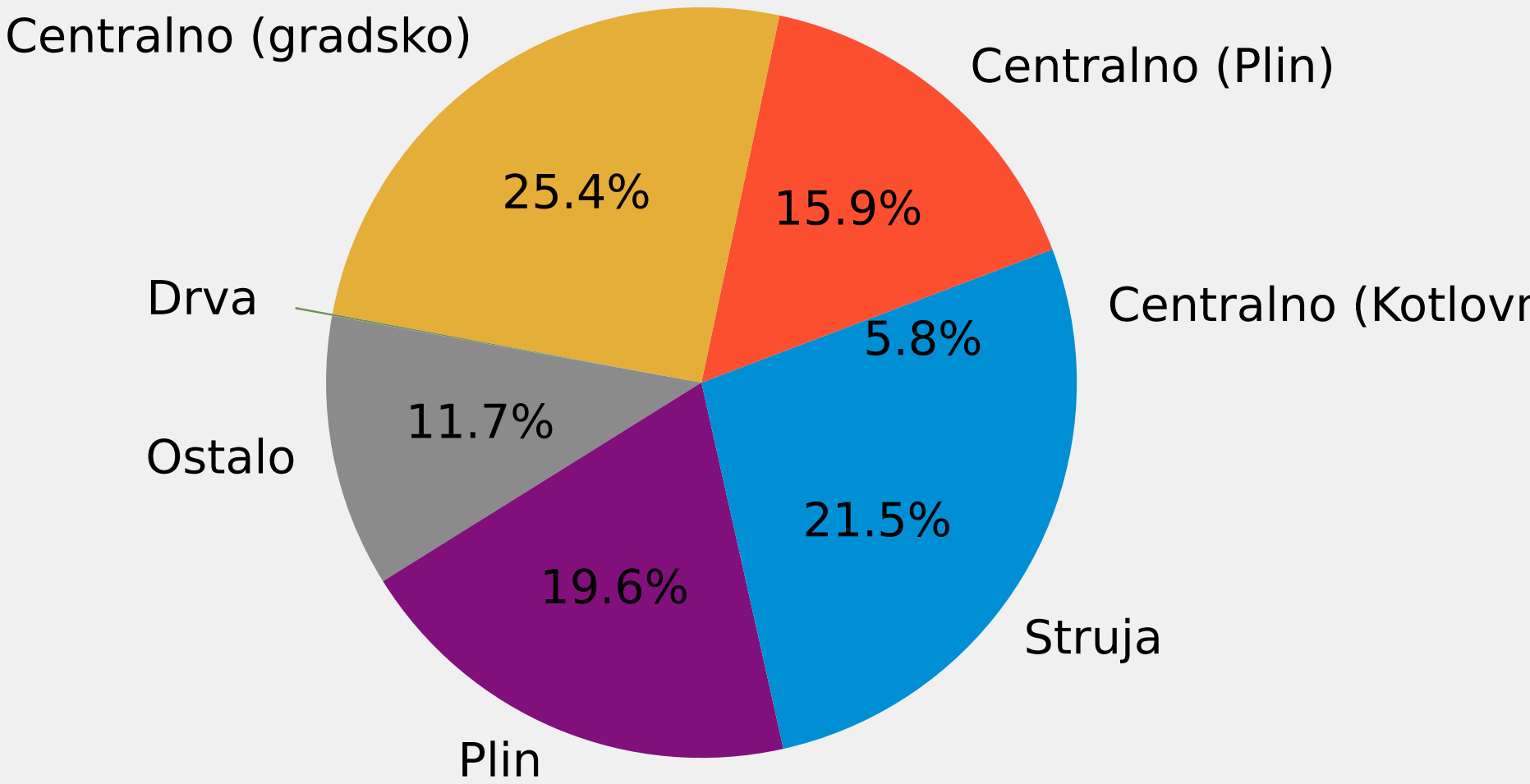


Heating - Rentals (Bar)

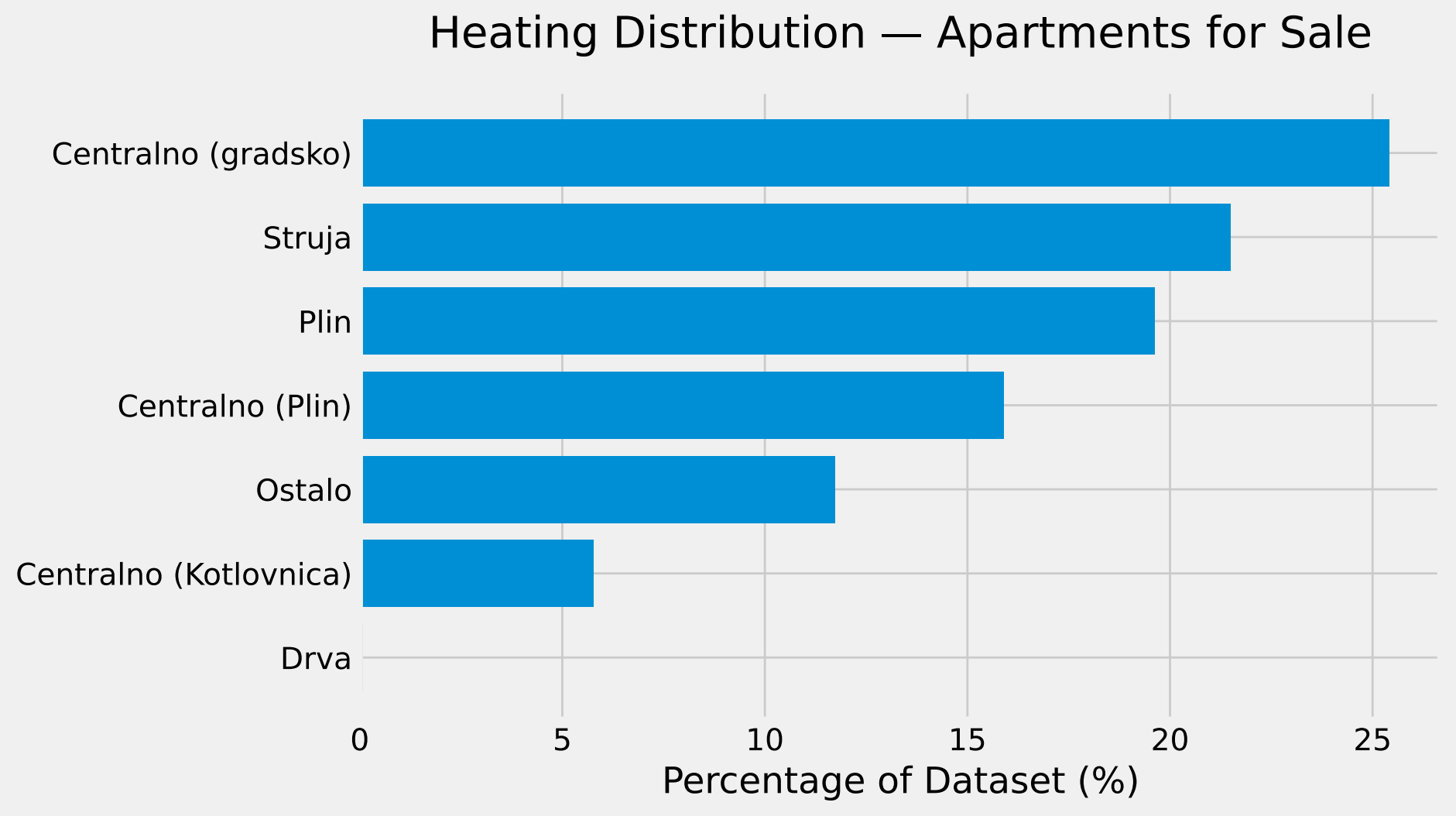


Heating - For Sale (Pie)

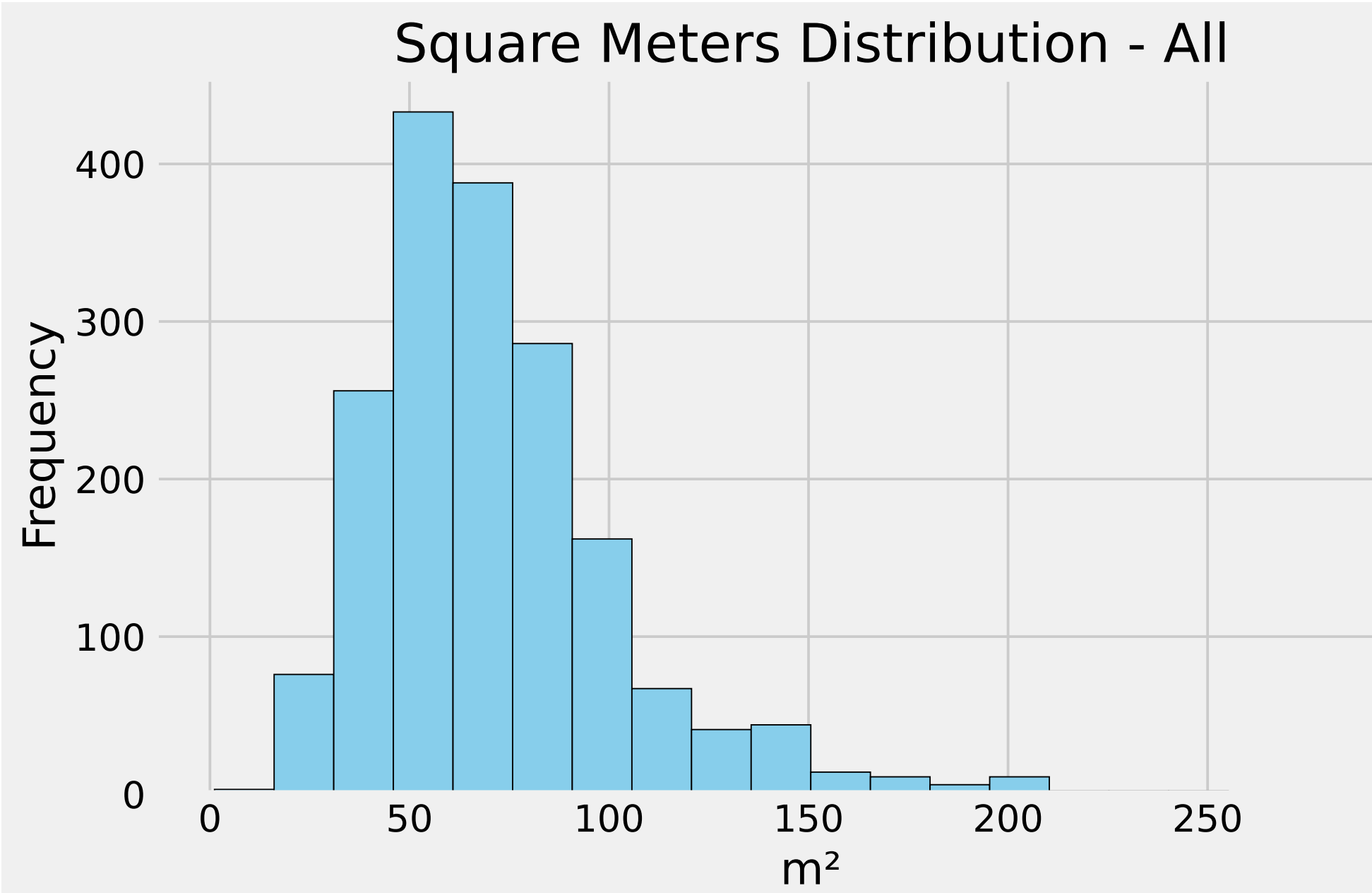
Heating Distribution for Apartments on Sale



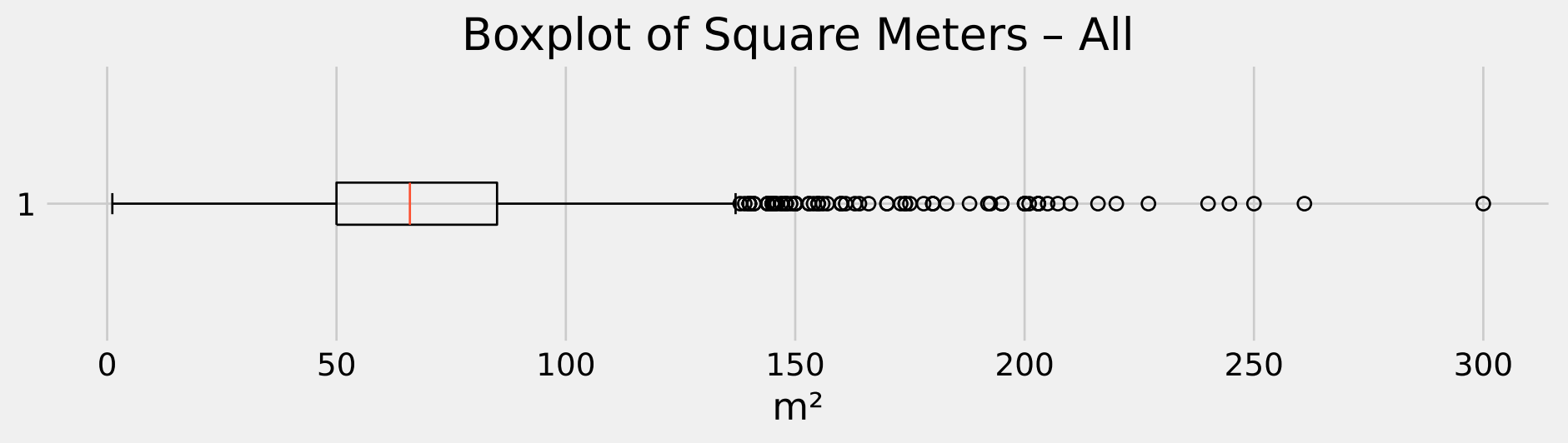
Heating - For Sale (Bar)



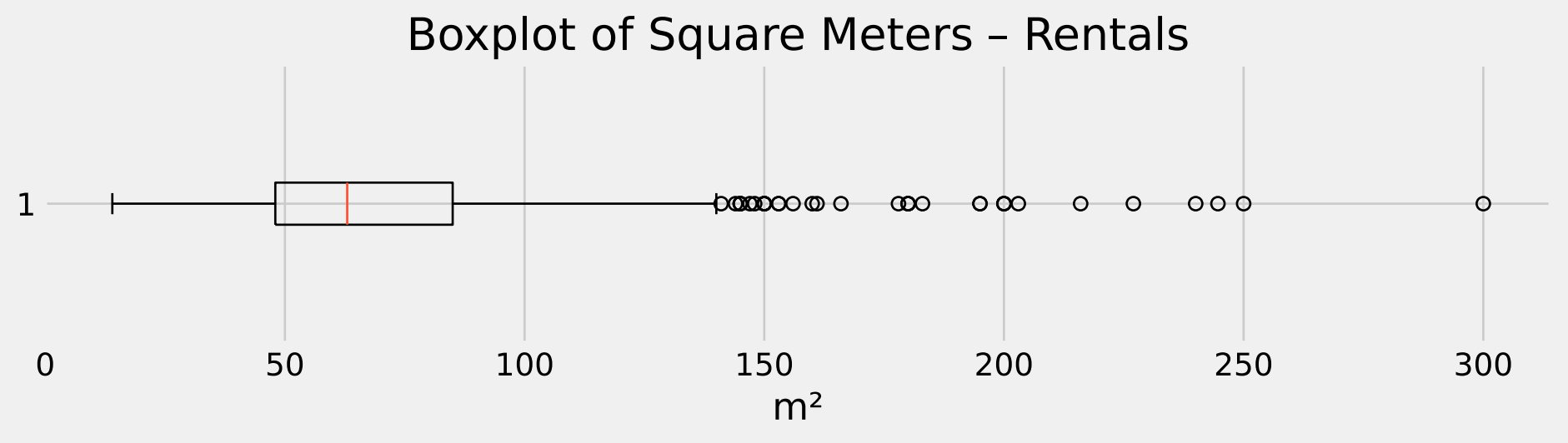
Square Meters - Histogram



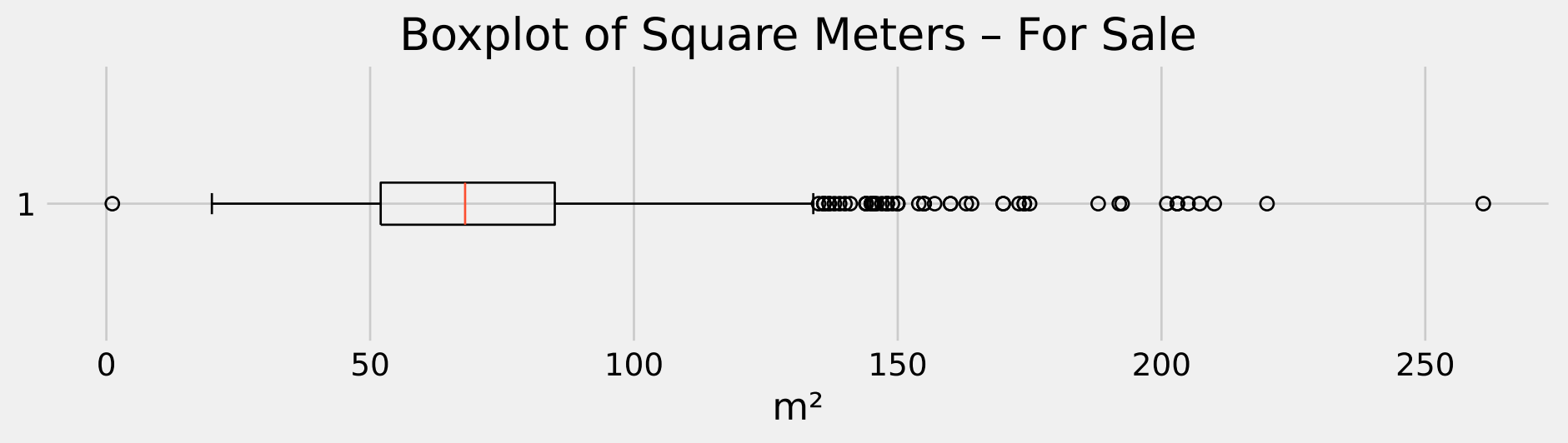
Square Meters - Boxplot (All)



Square Meters - Boxplot (Rentals)



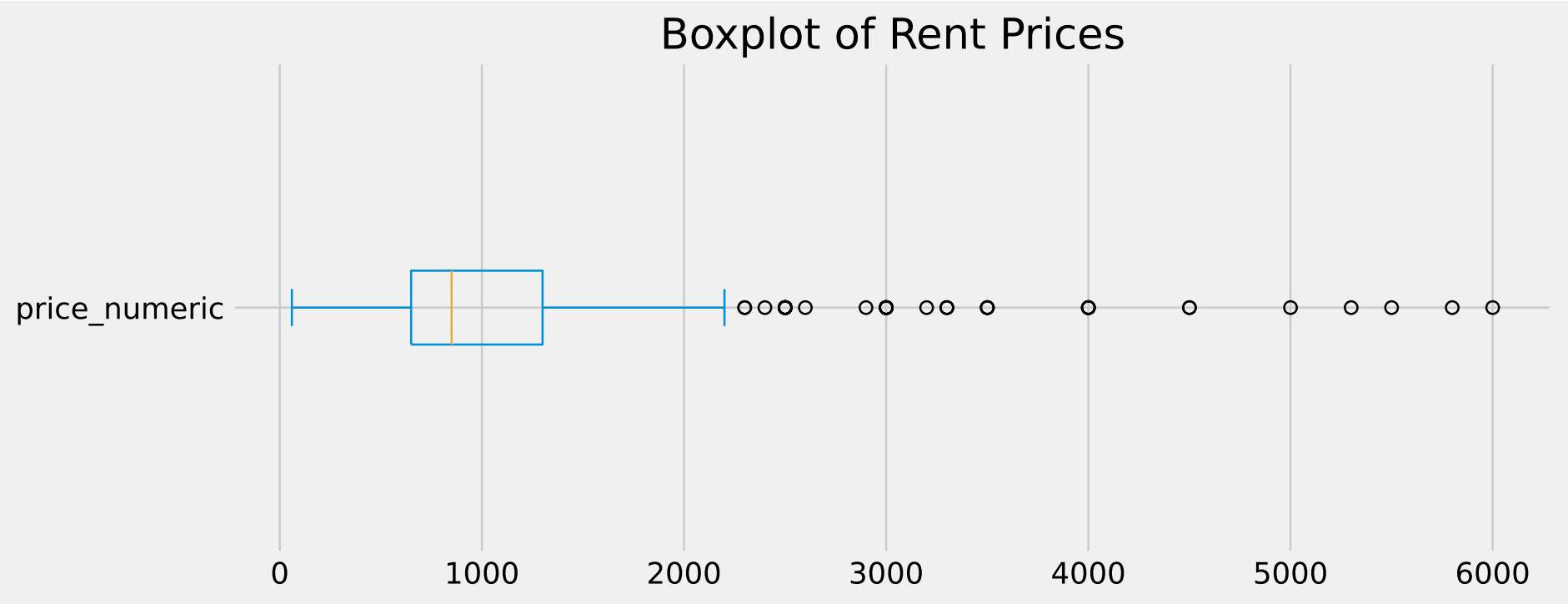
Square Meters - Boxplot (For Sale)



Square Meter Statistics

- **All Apartments** – Min: 1.13, Q1: 50, Median: 66, Q3: 85, Max: 300
- **Rentals** – Min: 14, Q1: 48, Median: 63, Q3: 85, Max: 300
- **For Sale** – Min: 1.13, Q1: 52, Median: 68, Q3: 85, Max: 261

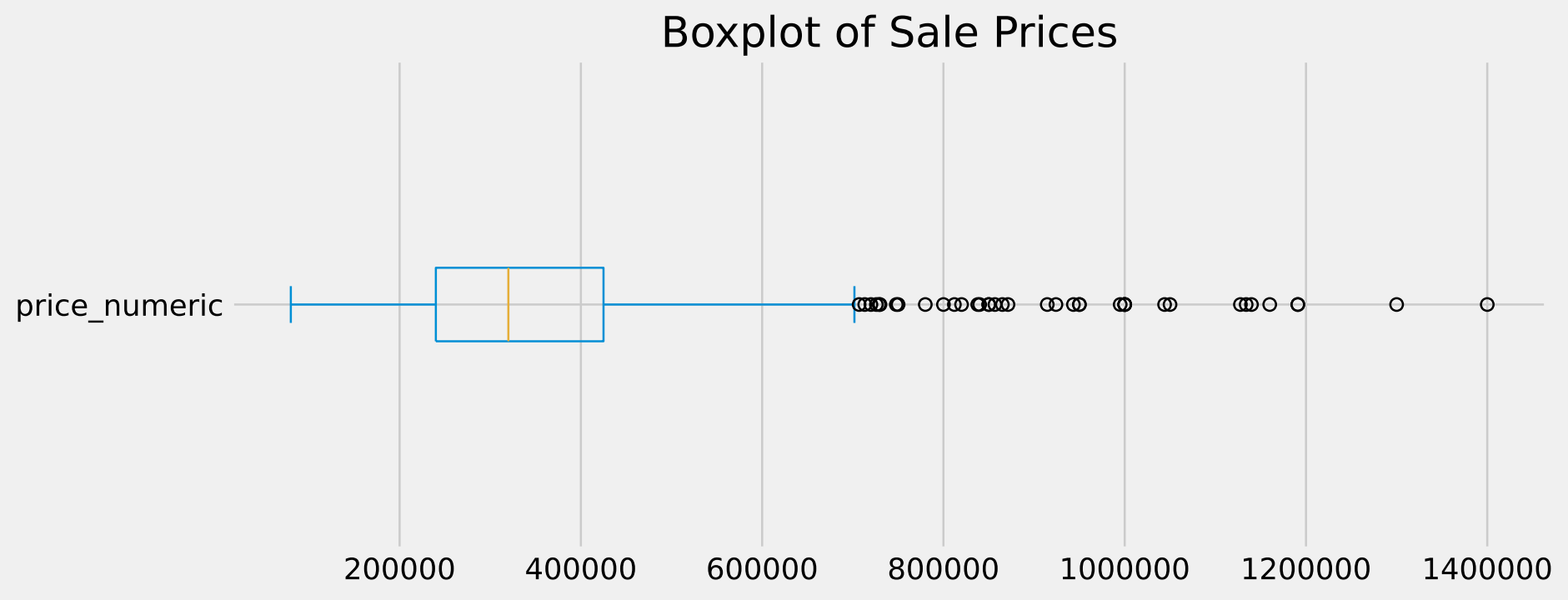
Rent Price - Boxplot



Rent Price Stats

- Min: 60 BAM, Q1: 650, Median: 850, Q3: 1300, Max: 6000
- Mean: 1153.66 BAM, Std Dev: 849.46

Sale Price - Boxplot



Sale Price Stats

- Min: 80,000 BAM, Q1: 240,000, Median: 320,000, Q3: 424,975, Max: 1,400,000
- Mean: 357,497 BAM, Std Dev: 177,897

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

Observed Limitations

- Some categorical features (e.g. `equipment`, `condition`, `heating`) have classes with very low representation.
  - These can introduce instability during training or bias metrics.
  - We will decide whether to merge rare classes or drop them depending on the context.
- **Most importantly:** the `municipality` feature is a poor indicator of location-based price variation.

Why Municipality is Not Sufficient

- For example, **Novi Grad** covers both:
  - high-demand areas near city centers,
  - and peripheral zones with low real estate value.
- Two apartments within Novi Grad may differ drastically in price due to local access to infrastructure and amenities.
- Hence, treating `municipality` as a proxy for "location" leads to poor model generalization.

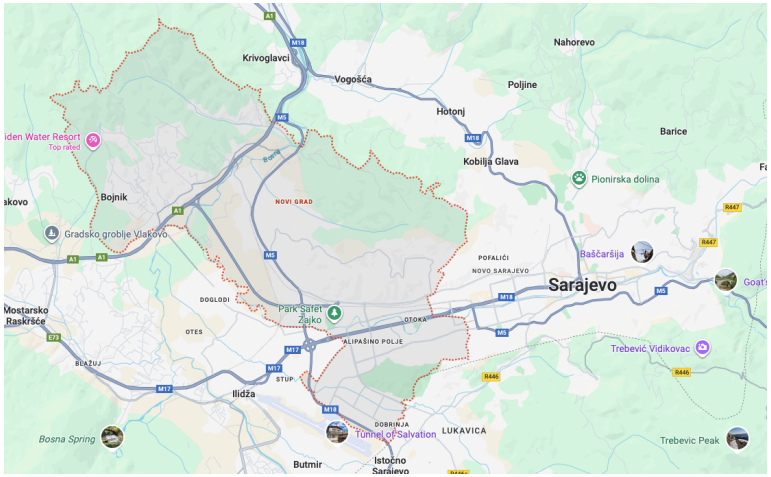


Illustration: Disparity in spatial price variation within one municipality.

Table of Contents

1. Introduction & Problem Statement
   - What is the real-world problem?
   - Why is it important or relevant?
   - Our Goal
2. Related Work
   - What has already been explored?
   - What do we offer that is new?
3. Data
   - Collecting The Data
   - Closer look at the features
   - Analysing the collected data
   - Analysing the Collected Data
   - Possible Issues with the data
   - Possible Solutions

Proposed Solution: Geo-based Enrichment

- We plan to extract the **longitude and latitude** of each apartment (either from address parsing or manual geocoding).
- With this data, we can:
  - Use the **Overpass API** and **OpenStreetMap** to query surrounding infrastructure.
  - Compute features like:
    - Number of tram/bus stations within 500m.
    - Distance to nearest hospital, park, supermarket, etc.
    - Average distance to n key public service locations.
- This makes our model more **location-aware** and enables predictions even in previously unseen areas.