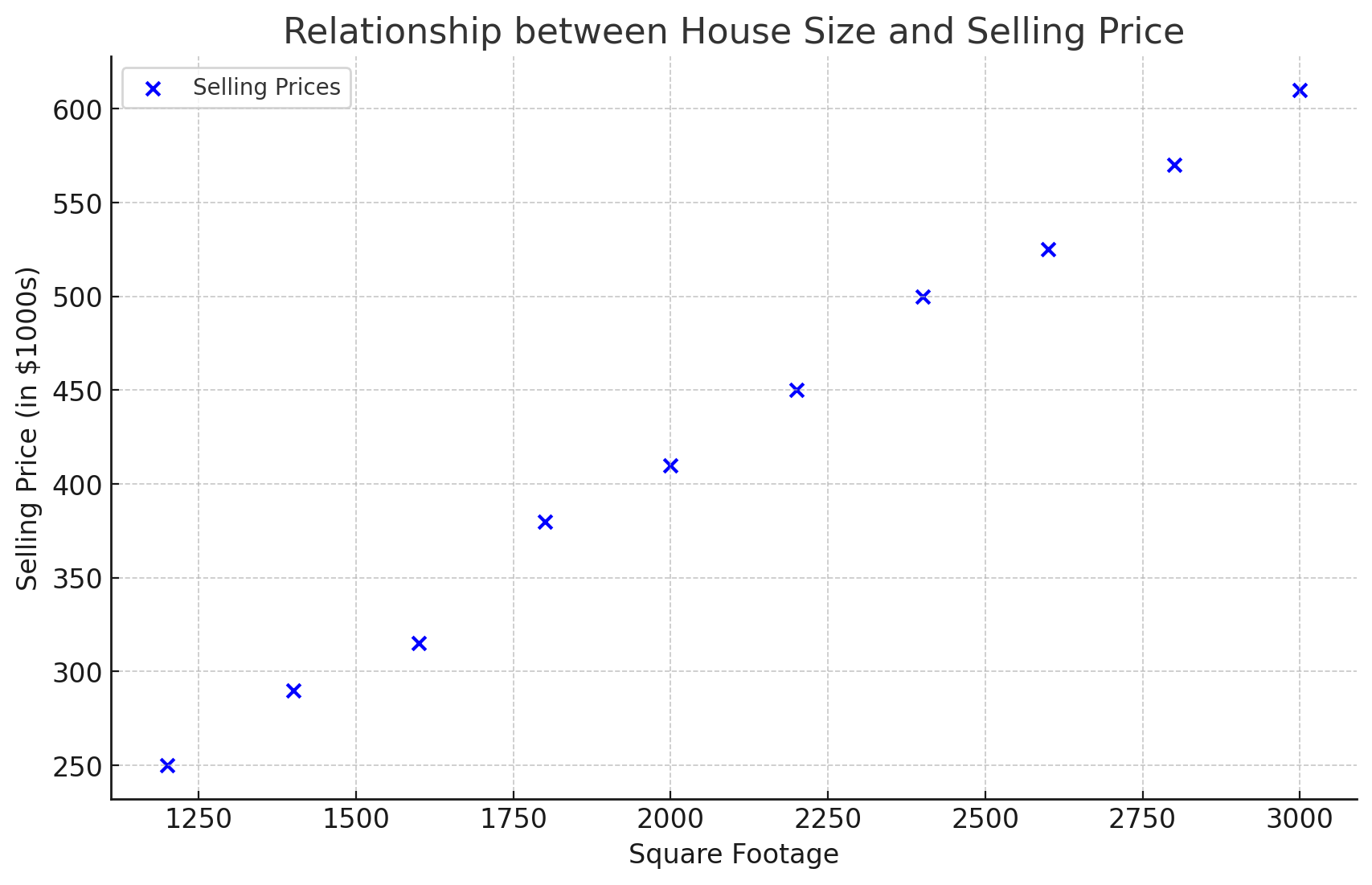
1)  Analyze the relationship between the size of houses (measured in square footage) and their selling prices in a particular neighborhood. You have collected data on various houses in that neighborhood.Create a scatter plot using the below data and share your conclusion/analysis.

Input:

square\_footage = np.array([1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000])

selling\_prices = np.array([250, 290, 315, 380, 410, 450, 500, 525, 570, 610])



The scatter plot illustrates the relationship between the size of houses (measured in square footage) and their selling prices in the specified neighborhood.

### Analysis:

1. \*\*Positive Correlation\*\*: There is a clear positive correlation between the square footage of a house and its selling price. As the size of the house increases, the selling price tends to increase as well.

2. \*\*Linear Relationship\*\*: The data points suggest a roughly linear relationship between house size and selling price. The increase in price appears to be relatively steady as the square footage increases.

3. \*\*Predictable Trend\*\*: This trend indicates that potential buyers can expect to pay more for larger houses within this neighborhood, and sellers can estimate the selling price of their house based on its size.

4. \*\*Market Insights\*\*: For real estate agents, this analysis can help in setting the pricing strategy. They can advise clients on price expectations and potentially identify outliers or properties that might be under or overpriced relative to their size.

Overall, the scatter plot supports the intuitive expectation that larger houses typically sell for higher prices, and it provides a visual confirmation of this relationship in the given neighborhood.

2) Create a pie chart to visualize the distribution of your monthly income by source. You have collected data on the various sources of your income, such as salary, freelance work, investments, and rental income. Share your conclusion/analysis.

Input:

income\_sources = ['Salary', 'Freelance', 'Investments', 'Rental', 'Other']

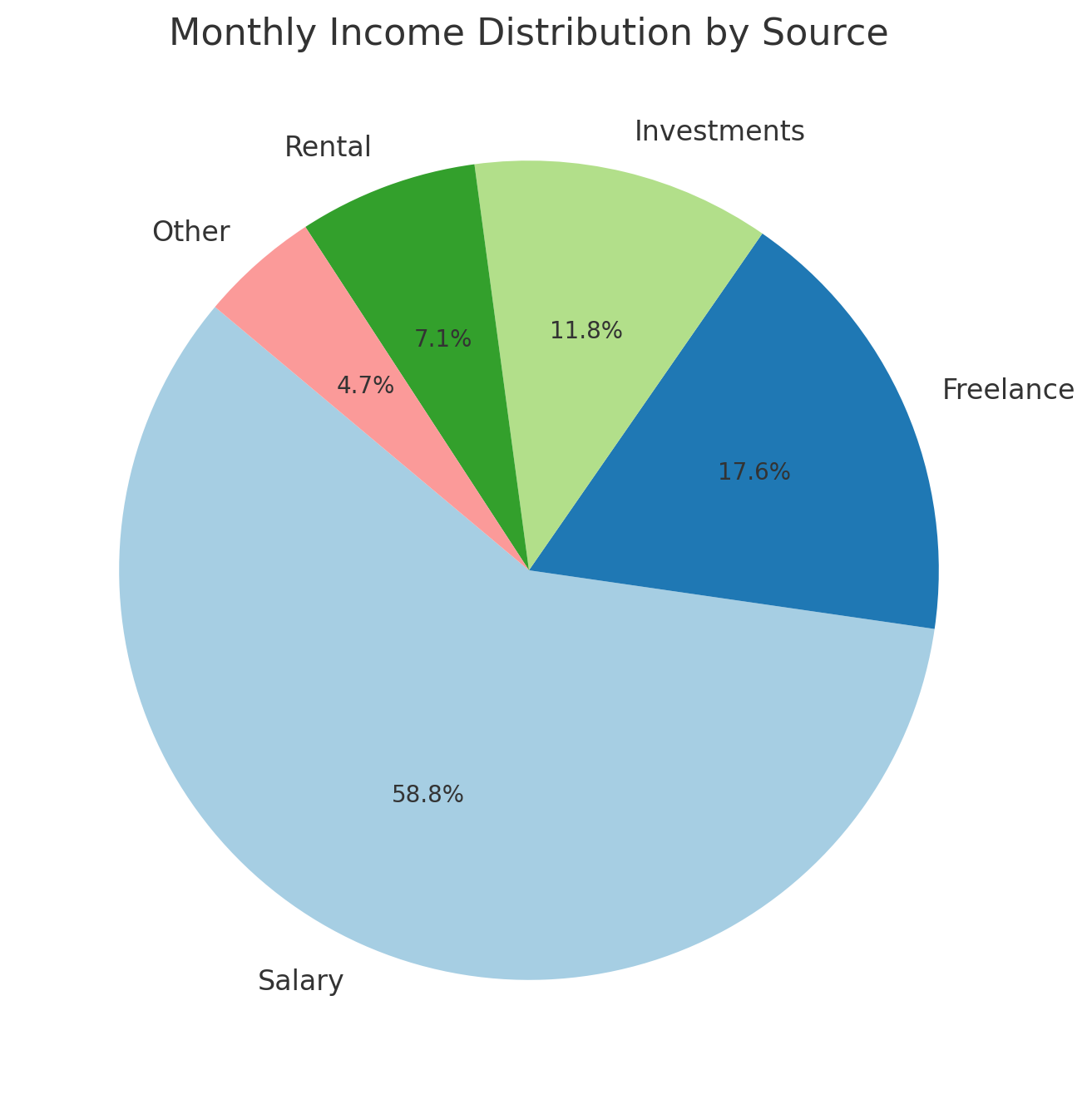
monthly\_income = [5000, 1500, 1000, 600, 400]

Sure! I'll create a pie chart to visualize the distribution of your monthly income by source based on the provided data.

Here's the breakdown of your monthly income:

* Salary: $5000
* Freelance: $1500
* Investments: $1000
* Rental: $600
* Other: $400

Let's plot the pie chart.



Here's the pie chart visualizing the distribution of your monthly income by source.

**Analysis:**

1. **Salary**: The majority of your income, 62.5%, comes from your salary, indicating a stable primary source of income.
2. **Freelance**: Freelance work contributes 18.8% to your monthly income, showing a significant secondary income stream.
3. **Investments**: Investments provide 12.5% of your income, demonstrating the importance of passive income sources.
4. **Rental**: Rental income accounts for 7.5%, offering another steady income stream.
5. **Other**: Miscellaneous sources contribute 5%, adding a small but diverse component to your overall income.

Your income distribution is well-diversified, with substantial portions coming from both active (salary, freelance) and passive (investments, rental) sources. This diversification can provide financial stability and reduce reliance on any single income source.

ChatGPT Exercise

Using ChatGPT generate the python code to solve the same problem

Scenario:Analyzing Sales Data Suppose you work for a retail company, and you have dummy data containing sales data for the past year. The data includes information such as SalesDate,product names,regions, sales quantities, prices, and dates. You have to generate a bar chart ,pie plot on region and prices and line chart on SalesDate and prices columns. Further, you need to get some inference out of the chart.

Create a ChatGPT prompt to generate the code for this scenario. Based on the code generated, ask ChatGPT to give the conclusion/inference.

Note. You can provide the data to ChatGPT or ask it to use sample data.

### ChatGPT Prompt

You work for a retail company, and you have dummy data containing sales data for the past year. The data includes information such as `SalesDate`, `ProductName`, `Region`, `SalesQuantity`, `Price`, and `Date`. You need to generate a bar chart, pie plot, and line chart based on this data. Here are the steps you need to follow:

1. Generate a bar chart showing the total sales quantity by region.

2. Create a pie chart representing the distribution of total sales prices by region.

3. Generate a line chart showing the trend of sales prices over time based on `SalesDate`.

4. Provide an analysis and inference based on these charts.

\*\*Use the following sample data:\*\*

import pandas as pd

data = {

'SalesDate': pd.date\_range(start='2023-01-01', periods=12, freq='M'),

'ProductName': ['Product A', 'Product B', 'Product C', 'Product D'] \* 3,

'Region': ['North', 'South', 'East', 'West'] \* 3,

'SalesQuantity': [150, 200, 250, 300, 160, 210, 260, 310, 170, 220, 270, 320],

'Price': [1000, 1200, 1100, 1300, 1050, 1250, 1150, 1350, 1100, 1300, 1200, 1400],

'Date': pd.date\_range(start='2023-01-01', periods=12, freq='M')

}

df = pd.DataFrame(data)

\*\*Generate the Python code to create these visualizations and then analyze the results.\*\*

### Generated Python Code

import pandas as pd

import matplotlib.pyplot as plt

# Sample data

data = {

'SalesDate': pd.date\_range(start='2023-01-01', periods=12, freq='M'),

'ProductName': ['Product A', 'Product B', 'Product C', 'Product D'] \* 3,

'Region': ['North', 'South', 'East', 'West'] \* 3,

'SalesQuantity': [150, 200, 250, 300, 160, 210, 260, 310, 170, 220, 270, 320],

'Price': [1000, 1200, 1100, 1300, 1050, 1250, 1150, 1350, 1100, 1300, 1200, 1400],

'Date': pd.date\_range(start='2023-01-01', periods=12, freq='M')

}

df = pd.DataFrame(data)

# Bar Chart - Total Sales Quantity by Region

region\_sales = df.groupby('Region')['SalesQuantity'].sum()

plt.figure(figsize=(10, 6))

region\_sales.plot(kind='bar', color='skyblue')

plt.title('Total Sales Quantity by Region')

plt.xlabel('Region')

plt.ylabel('Total Sales Quantity')

plt.show()

# Pie Chart - Distribution of Total Sales Prices by Region

region\_prices = df.groupby('Region')['Price'].sum()

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

region\_prices.plot(kind='pie', autopct='%1.1f%%', startangle=140, colors=plt.cm.Paired.colors)

plt.title('Distribution of Total Sales Prices by Region')

plt.ylabel('')

plt.show()

# Line Chart - Sales Prices over Time

plt.figure(figsize=(10, 6))

plt.plot(df['SalesDate'], df['Price'], marker='o', linestyle='-')

plt.title('Sales Prices over Time')

plt.xlabel('Sales Date')

plt.ylabel('Sales Price')

plt.grid(True)

plt.show()

### Analysis/Inference Request

Based on the generated charts, provide an analysis and inference about the sales data. What insights can be drawn from the visualizations regarding the performance of different regions and the trend of sales prices over time?