Individual Coursework Submission Form

Specialist Masters Programme

Surname: Rawat	**First Name: Nikhil**
MSc in: Business Analytics	**Student ID number: 230044335**
Module Code: SMM750	
Module Title: Digital Technologies and Value Creation	
Lecturer: Professor Philippe Blaettchen	**Submission Date: 04/19/2024**

Declaration:

By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work, I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.

Marker's Comments (if not being marked on-line):

Deduction for Late Submission:

Final Mark: %

Part 1: Preparing the Data

Combining and Pre-processing the Dataset

Given the dataset spanning monthly data from 2015 to 2021, I structured the pre-processing into well-defined steps to ensure data integrity:

1. **Data Segmentation**:
    - The dataset was divided into Data_A (2015-2019) and Data_B (2020-2021) due to discrepancies in data format, including a broader historical range in Data_B from 1963 to 2021.
2. **Standardization**
    - **Column Standardization**: Utilized a dictionary to map and merge various column names into standardized names across both datasets.
    - **Start-up Column**: Removed empty rows in the start-up column that lacked critical information.
    - **Irrelevant Columns**: Removed columns with high missing values and those irrelevant to our analysis, such as Serial No., About Company, and Founder Name, which provided no added value.
3. **Date Handling and Data Integration:**
    - **Date Correction for Data_B:** Converted non-numeric date entries to integers. Dropped entries outside the 2020-2021 range, aligning the focus strictly on our period of interest. Standardized dates to DD/MM/YYYY format.
    - **Merging Datasets:** Combined Data_A and Data_B into a unified dataset after aligning date formats and ensuring consistency across data fields to facilitating a comprehensive analysis across the full date range.
4. **Missing Values Handling Strategy:**
    - **Investment Types**: Created and applied a dictionary to map diverse investment type entries to standardized categories. This strategy helps in ensuring that investment types are consistently labelled, which is crucial for analysing trends in funding types and comparing investment patterns over the years. Then missing values were filled using the top most prevalent investment type based on occurrence probability.
    - **Industry Verticals**: Missing values in this category were populated using the top two industries (Consumer Internet and Technology), reflecting their dominance in the dataset. Values were assigned based on their occurrence probability.
    - **Amounts**: Corrected non-numeric entries to NaN and converted the rest to numeric. To address missing values: **Industry-specific mean:** Imputed with the average amounts from similar industry verticals, Maintaining the relative scale and distribution of investments within industries. **General median:** For industries without enough data to compute a meaningful average, the median investment amount from the entire dataset was used. This approach mitigates the impact of outliers in highly skewed data and provides a more robust filler for missing values.
    - **Cities and Investors**: The missing values for city locations and investor names were filled based on the most frequent entries based on their occurrence probability, reflecting the concentration of activities in major cities and by prominent investors as in both only few were dominating with great margins.
5. **Data Export**: The fully processed and integrated Data Frame was then saved as a CSV file for further analysis or reporting.

Impact on Analysis: To ensure the integrity of my analysis, missing values within our dataset were systematically addressed. By mapping investment types to standard categories and populating these fields based on the most common type, I preserved the consistency essential for identifying investment trends. For the 'Amount(in USD)' column, I balanced industry-specific imputations with general median values to maintain a realistic scale of investment across sectors and to mitigate skew from outliers. Filling missing 'City' and 'Investor' names according to the frequency of occurrence allowed our dataset to accurately reflect the concentration of start-up activities and investor presence This process ensures that the dataset is not only comprehensive but also accurately represents the dynamics of the start-up ecosystem in India from 2015 to 2021.

Part 2: Strategic Positioning

Analysing the start-up landscape from 2015 to 2021(Please refer appendix for the figures) reveals key factors influencing funding:

1. **Yearly Trends**: The initial surge in the number of start-ups in 2015 and 2016, with a peak in 2016, suggests a booming interest in entrepreneurship, likely due to favourable policies such as "Make in India" policy or market conditions. However, the following decline especially could be attributed to market saturation, increased competition, or economic factors. (Fig. 1)
   Interestingly, the total funding amount peaked in 2017, diverging from the trend in start-up numbers. This peak reflect a few large funding rounds or a shift towards investment in scaling businesses rather than a multitude of new entrants. A strategic takeaway for start-ups is the importance of timing market entry and capitalizing on favourable conditions. (Fig. 2)
2. **Industry Preference**: Dominance in funding for Consumer Internet and Technology underscores the critical role of digital transformation and innovation(Fig. 3). Additionally, Sectors like Transportation and E-commerce platforms show substantial average funding, signalling investor confidence in these models' scalability and potential return on investment(Fig. 4). For a start-up seeking funding, positioning within these high-growth, high-investment sectors could be advantageous.
3. **Geographical Impact**: The geographic distribution of funded start-ups indicates a clear preference for certain cities, with Bangalore, Mumbai, and New Delhi leading(Fig. 5). This suggests that location may be a significant factor for start-ups looking to secure funding, potentially due to better ecosystems, access to investors, and networking opportunities in these cities. For start-ups not located in these hubs, it could be beneficial to establish a strong virtual presence or consider relocation.
4. **Investment Type Preferences**: The investment type graph indicates a preference for Seed and Equity funding, implying that early-stage start-ups have a good chance of securing initial rounds of funding, and there is ample growth funding available for scaling businesses. (Fig. 6)
5. **Distribution of Funding**: The distribution graph is heavily right-skewed (Fig. 7), which suggests that while there are many start-ups receiving smaller amounts of funding, a few outliers receive substantially larger investments. This skewness highlights the competitive nature of significant funding rounds and suggests that start-ups need to have a standout proposition to attract larger investments.

To position my start-up for investment, considering the analysed trends and data, I would:

1. **Capitalise on Timing**: Launch fundraising efforts when the market shows a high appetite for entrepreneurship, like during a surge in policy-driven investments or technology booms. I'd stay attuned to market conditions and policy landscapes to identify opportune moments for seeking investment.
2. **Align with Growth Sectors**: I'd steer my start-up into sectors like Consumer Internet and Technology, which command significant investor interest. I'd ensure my business model is aligned with the scalability and ROI that sectors like Transportation and E-Commerce platforms demonstrate, showcasing the potential for high growth.
3. **Harness Geographic Advantage**: If practical, I'd set up base in a start-up hub like Bangalore or Mumbai to access a better investor ecosystem. Otherwise, I'd create strong virtual connections to these hubs, maximizing networking opportunities and investor exposure.
4. **Seek Preferred Investment Types**: I'd tailor my pitch and business plan to attract Seed and Equity funding, which are prevalent in the ecosystem. For early-stage investment, I'd highlight the market potential and product-market fit, while for growth stages, I'd present a solid plan for scaling and profitability.
5. **Craft a Standout Proposition**: To compete for significant investment in a skewed distribution, my start-up would need a compelling value proposition. Whether it's through innovative solutions, a disruptive model, or exceptional market traction, I'd ensure that my start-up's offerings are distinct and attractive to investors looking for high returns.

By embedding these strategic insights into my start-up's approach—from timing and sector positioning to geographic presence and investment targeting—I'd substantially improve my start-up's prospects for securing investment.

Appendix:

Figure 1: Number of Start-ups per Year

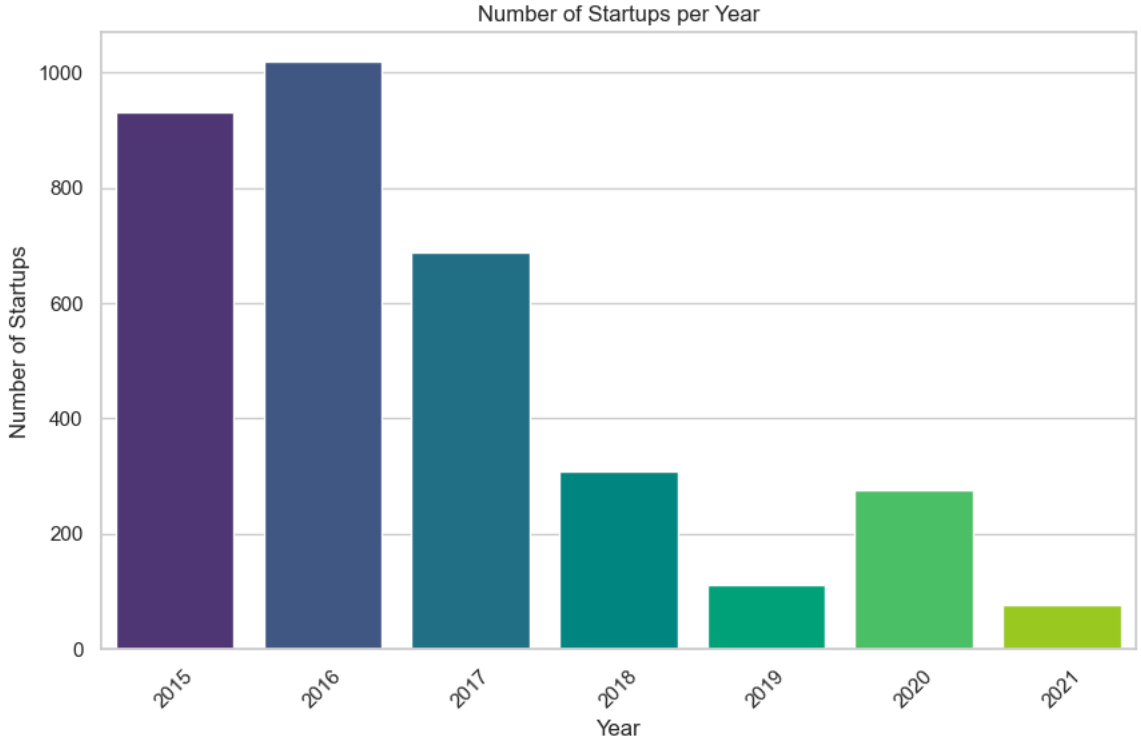

Figure 2: Total Funding Amount Over Years

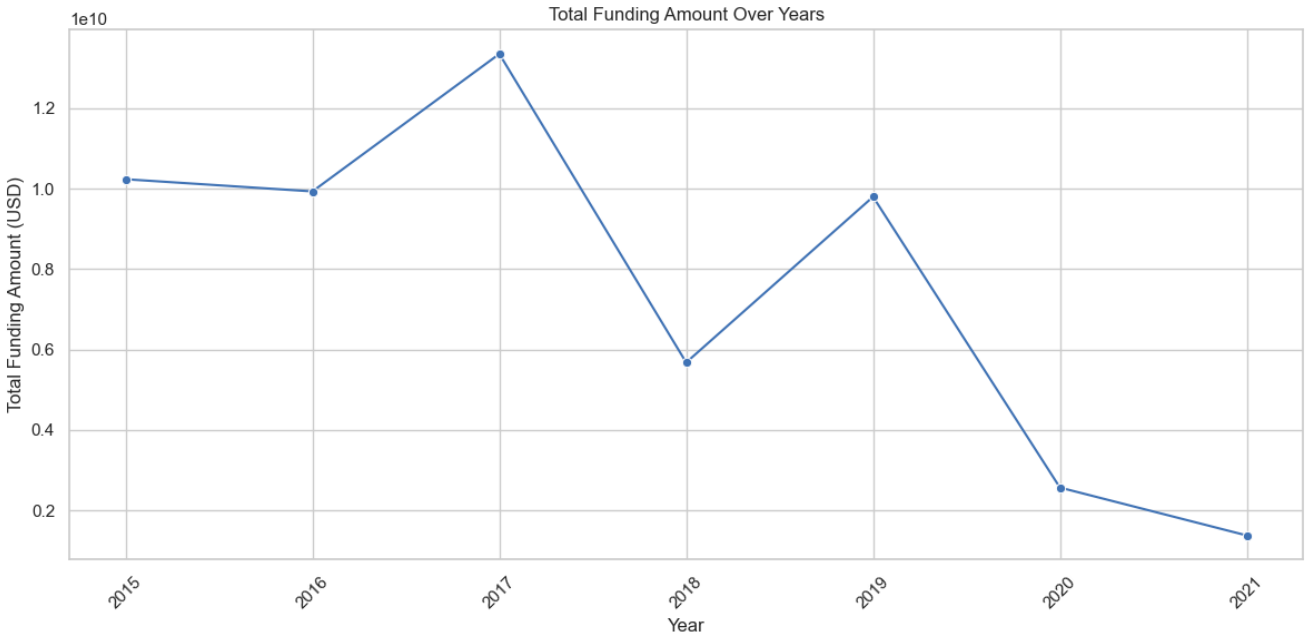

Figure 3: Distribution of start-up counts by Industry vertical

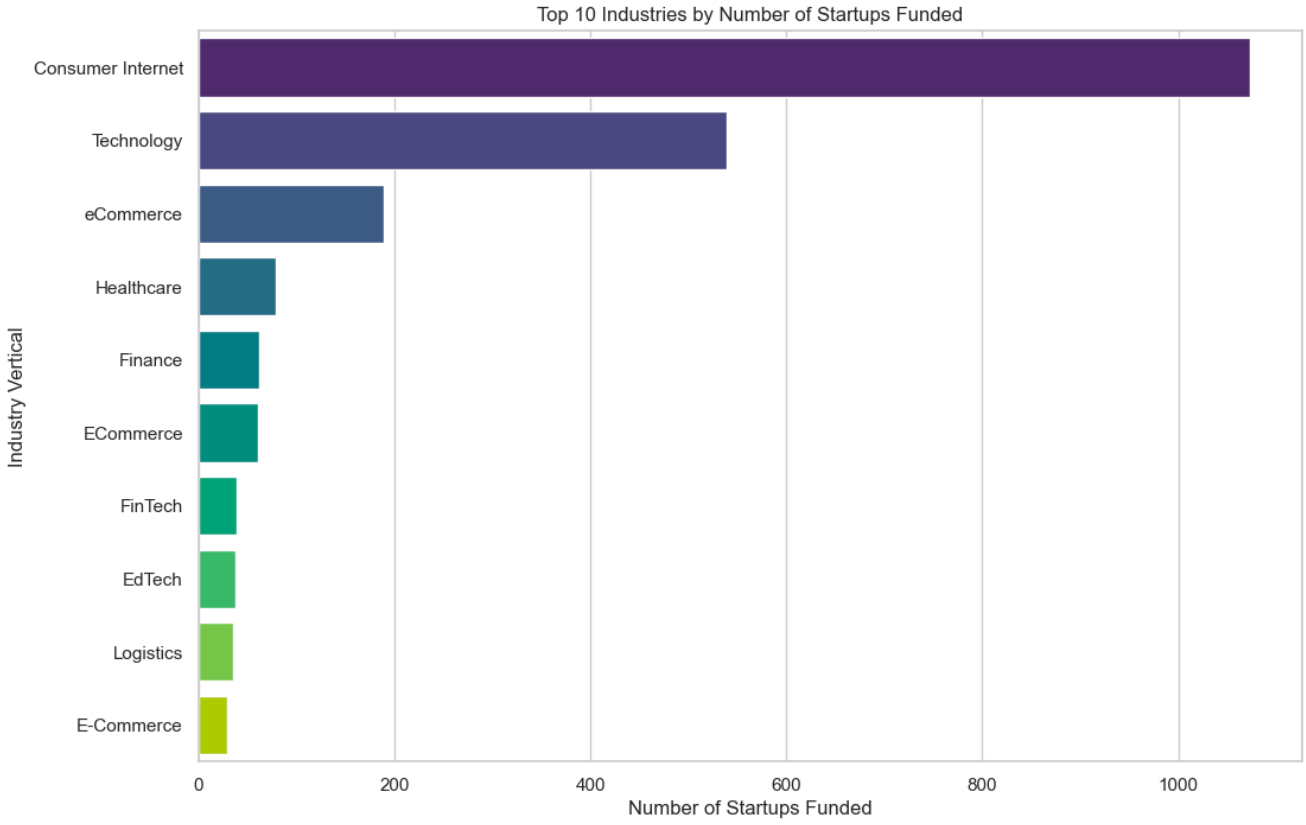

Figure 4: Average Funding Amount by Industry Vertical

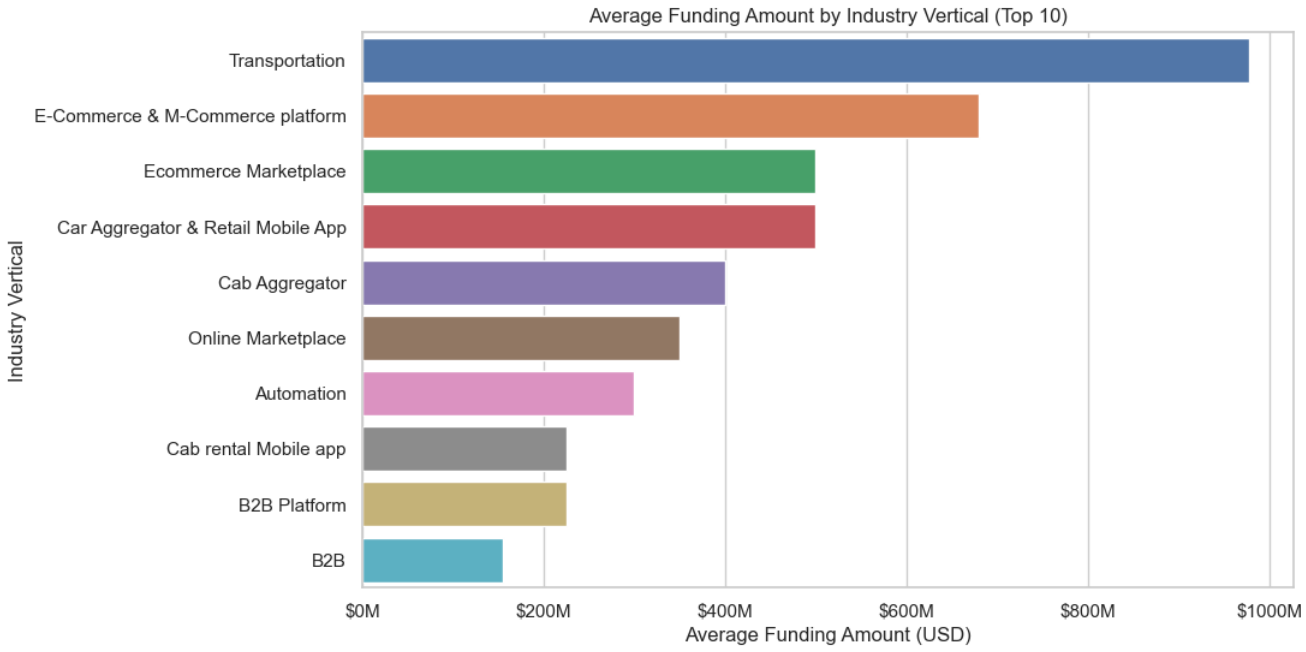

Figure 5: Distribution of start-up counts by City location

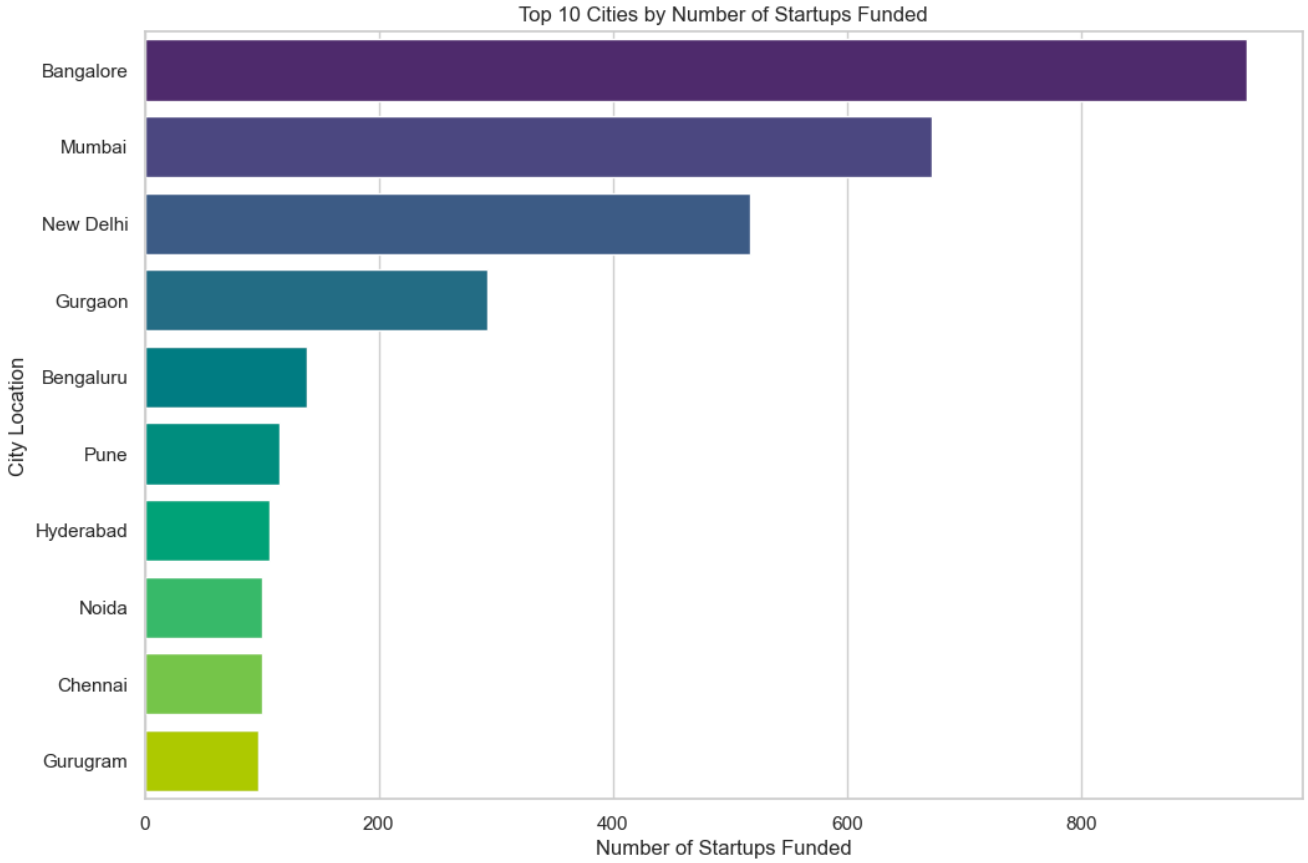

Figure 6: Investment Type analysis

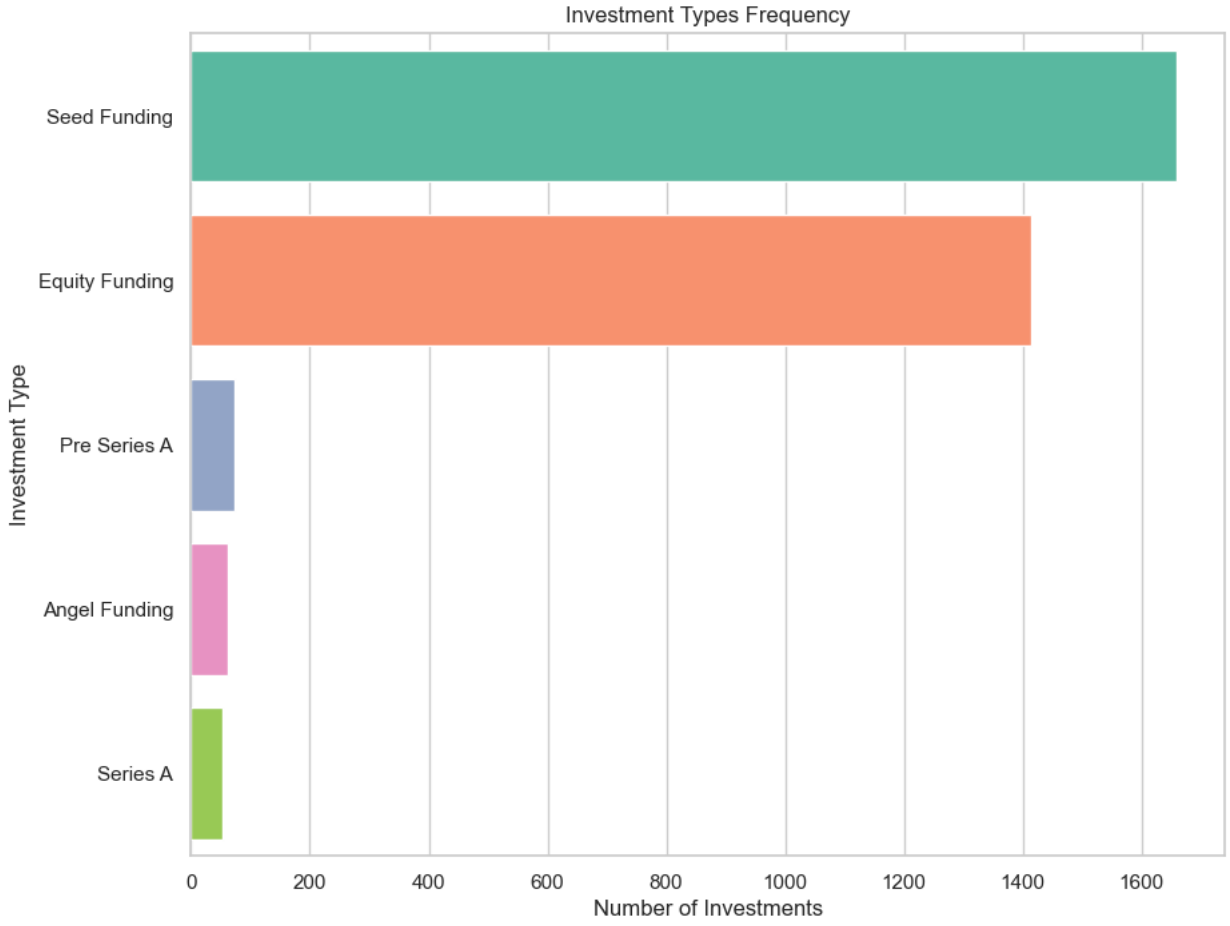

Figure 7: The distribution of funding amounts