IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Data Plotting And Visualization Using Openai

[1]Kaushik NS, [2]Mr.K. Santhosh Kumar, [3]Mr.Yuvraj Sharma
[1]PG Student, [2]Assistant Professor, [3]Chief Technical Officer
[1,2]School of Management,
[1,2]Hindustan University, Chennai, India
[3]PlatypusBox, Delhi, India

Abstract:

Background: The intersection of artificial intelligence and data visualization opens new opportunities for automation, accessibility and enhanced user experience.
Objectives: This study aims to leverage the capabilities of OpenAI, specifically its generative AI models, to create advanced and
dynamic data plotting and visualization tools.
Methodology: Video/Online Games related primary data will be taken from Steam (Steam is one of the most popular digital video game distribution platforms to buy and play videos games). Steam game related datasets will be taken from Kaggle (Kaggle is a platform for data science and machine learning professionals and offers public data sets) or through a API program. The data will be transformed to visualized representations using OpenAI to provide detailed insights into the games and the gaming market.
Results: The project will result in a tool or proof of concept that combines OpenAI-driven insights with advanced data plotting and
visualization techniques. This will enhance data exploration, making it more intuitive and interactive for users of all levels.
Conclusion: By merging AI with data visualization, this project seeks to simplify complex data interpretation and enable deeper
insights for users through intuitive, visually rich interfaces powered by OpenAI.

This study has been undertaken to investigate the determinants of stock returns in Karachi Stock Exchange (KSE) using two assets pricing models the classical Capital Asset Pricing Model and Arbitrage Pricing Theory model. To test the CAPM market return is used and macroeconomic variables are used to test the APT. The macroeconomic variables include inflation, oil prices, interest rate and exchange rate. For the very purpose monthly time series data has been arranged from Jan 2010 to Dec 2014. The analytical framework contains**.**

Keywords: **OpenAI, Data Visualisation, Video Games, AI, ChatGPT**

I. INTRODUCTION

The data visualization and analytics industry plays a critical role in transforming raw data into meaningful insights for businesses, governments, and researchers. With the growing volume of data generated globally, the demand for effective visualization tools has surged across various sectors, including finance, healthcare, education, and marketing.

OpenAI is a leading research organization and technology company focused on advancing artificial intelligence (AI) in a safe and beneficial manner. Founded in December 2015, OpenAI aims to ensure that artificial general intelligence (AGI)—highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity

Data visualization is the graphical representation of information and data. By using visual elements such as charts, graphs, maps, and other formats, data visualization helps to transform complex datasets into easily digestible insights. This field combines the principles of data analysis with visual design to enhance understanding and facilitate decision-making.

Data visualization is a cornerstone of decision-making and communication in numerous fields, ranging from business analytics to scientific research. The integration of AI, particularly OpenAI technologies, into data visualization opens new possibilities for enhancing the way data is analyzed, interpreted, and communicated

- This study helps us to know how OpenAI's natural language processing capabilities can enable users to generate visualizations through conversational queries, making data analysis accessible to a broader audience
- This study helps us to know how OpenAI can automate tasks such as data preprocessing, selecting visualization types, and generating plots, significantly improving efficiency.

II. DATA COLLECTION

2.1 DataSet

The Steam Games Dataset on Kaggle, provides a comprehensive collection of data about games on the Steam platform. It includes a variety of information for over 27,000 games, offering insight into aspects like game pricing, ratings, genres, and release dates. This dataset is especially valuable for analysis in game development, market research, and trend prediction within the gaming industry.

Below are the breakdown of key information available in the dataset:

- **Game Information**: Includes game titles, developers, publishers, and release dates, which can help in analyzing trends over time or studying developer-specific data.
- **Pricing and Sales**: Details about game prices, discounts, and overall popularity (such as owner estimations) are included, allowing for price and sales analysis.
- **Genre and Categories**: Games are categorized by genre, such as Action, Adventure, Indie, and more, and contain tags like "Multiplayer" or "Single-player," useful for genre-based studies and feature exploration.
- **Ratings, Reviews & Playtime**: The dataset contains user review counts and ratings, which can be used to assess a game's popularity and reception over time. It also provides details on the playtime by the registered player.
- **Miscellaneous**: It also includes data like supported platforms (Windows, macOS, Linux), which can be helpful in market segmentation and cross-platform analysis.

This dataset is well-suited for machine learning, such as predicting game popularity, analysing pricing strategies, or identifying genre trends over time. It can be used for building recommendation systems, studying consumer behaviour, and exploring the dynamics of the gaming industry.

The collected data is for all PC Games published on steam from 1997 to May 2024. (across 27 years) which contains 83,656 items

2.2 Data source

Steam Gaming Platform	• Steam is a platform and app for distributing video games online that also has social media features and community message boards • There are over 73 thousand games available on Steam, with over nine thousand games being published in the first nine months of 2023 alone • Steam collects a variety of data from its users to enhance their gaming experience, improve services, and facilitate community interactions.
Kaggle data science community	• Kaggle is a popular platform designed for data science and machine learning enthusiasts.

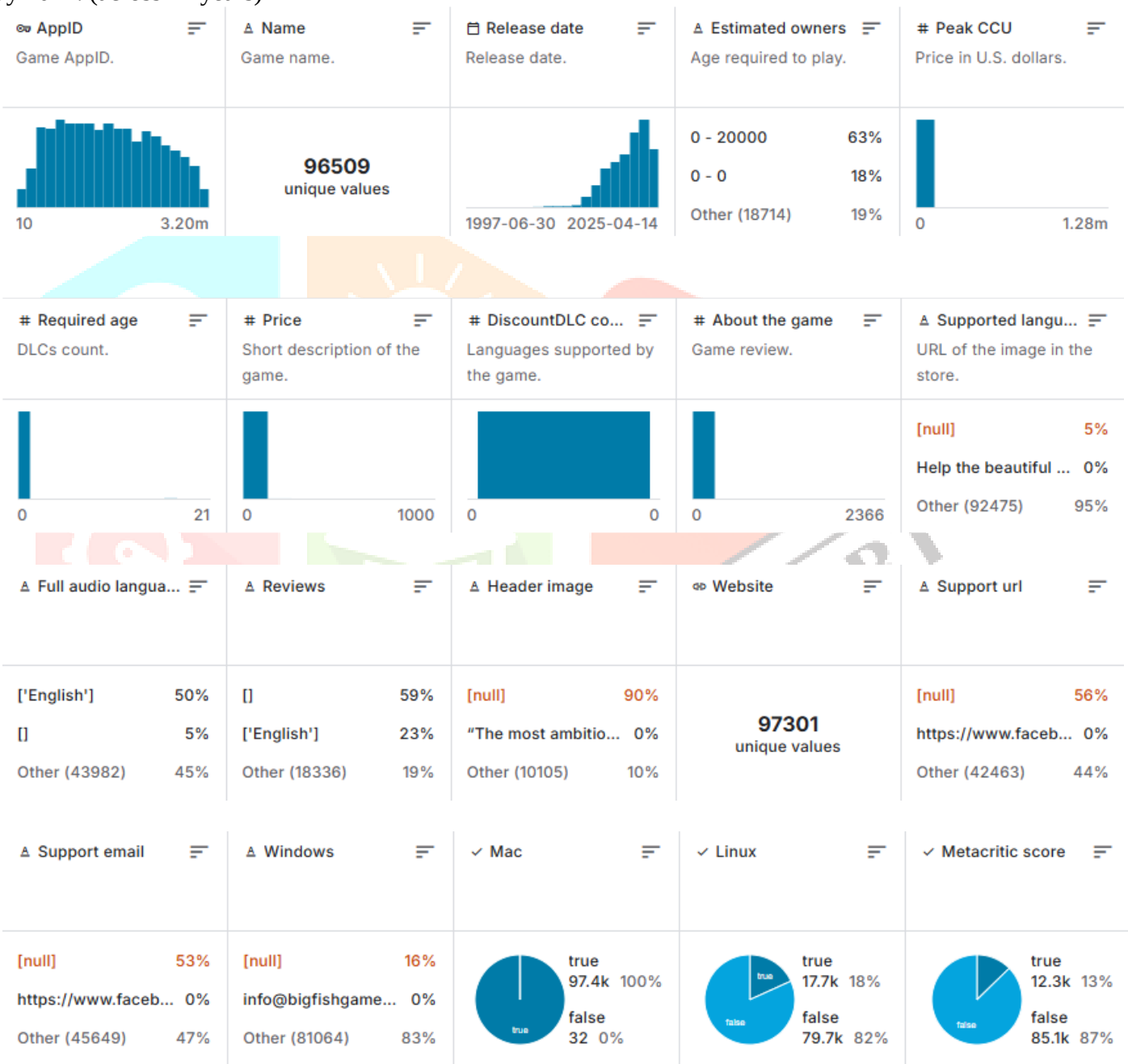
- Founded in 2010 and acquired by Google in 2017, it offers a rich community experience where users can engage in data science projects, enhance their skills, and collaborate with others.

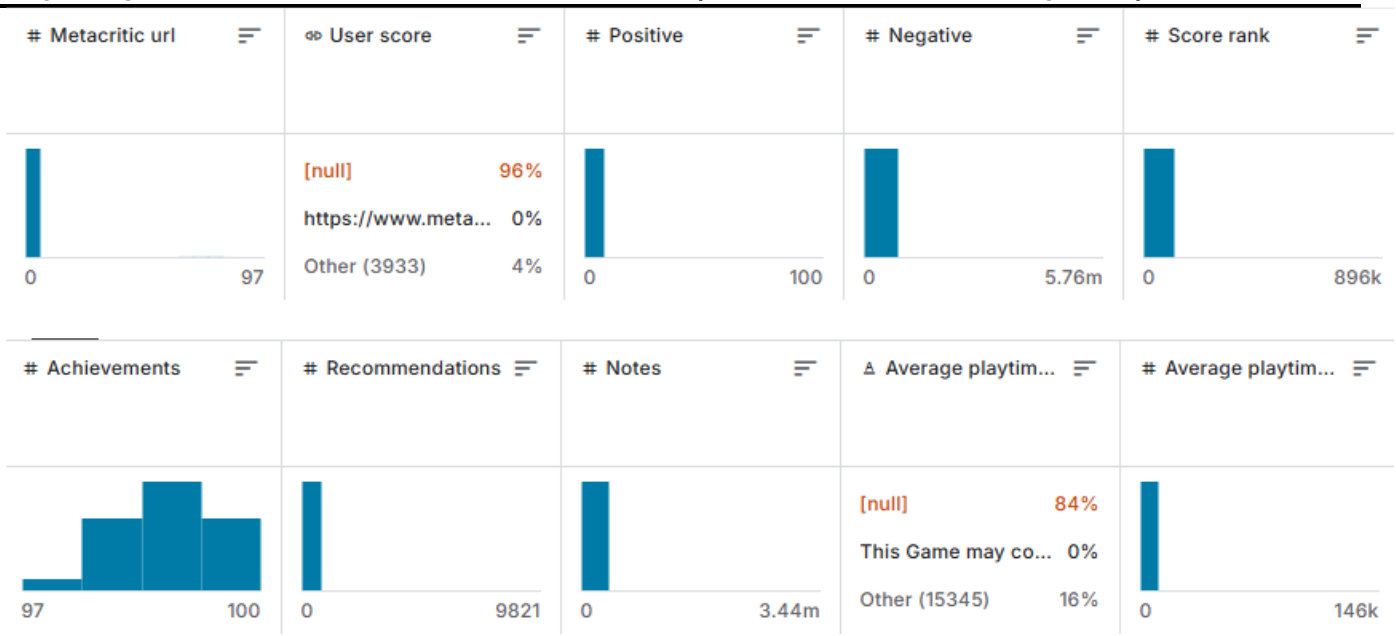
III. Result & Discussion

3.1 Analysis of Data

The data analysis focused on exploring the potential of OpenAI for data visualization tasks. By using OpenAI to process, summarize, and interpret datasets, the project identified key patterns and trends.

Below are the key data set details were used which is for all PC Games published on steam from 1997 to May 2024. (across 27 years)





The data analysis focused on exploring the potential of OpenAI for data visualization tasks. By using OpenAI to process, summarize, and interpret datasets, the project identified key patterns and trends

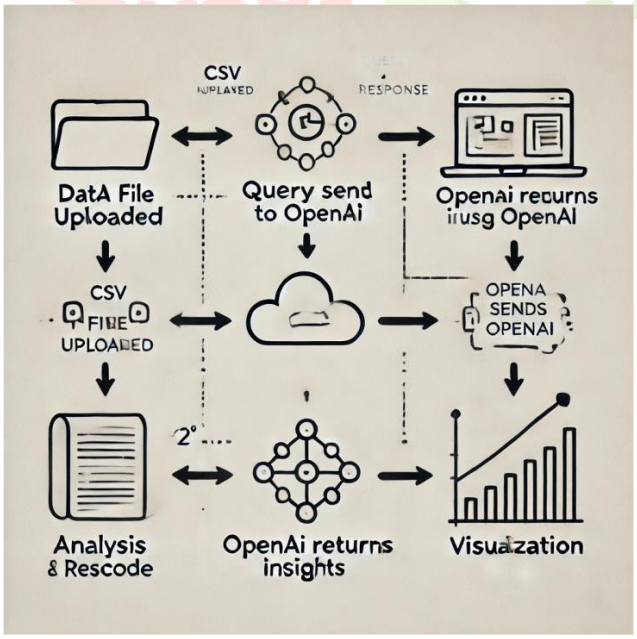
3.2 Tools for Analysis

Python 3.12.7	Python 3.12 is the newest major release of the Python programming language, and it contains many new features and optimizations. 3.12.7 is the latest maintenance release, containing about 120 bug fixes, build improvements and documentation changes since 3.12.6. Plenty of tools and languages are available for data visualization, but Python is an ideal choice because it has a small line of code. Not just that, it has an easy syntax and takes less time to code compared to other languages. Another plus point is that Python has numerous libraries and packages for data visualization, making it easy to quickly create visual data.
OpenAI	OpenAI is an artificial intelligence research laboratory. Established in December 2015, OpenAI is dedicated to advancing artificial general intelligence (AGI) in a safe and beneficial manner. The organization was founded by a group of influential tech visionaries, including Sam Altman, Greg Brockman, Ilya Sutskever, John Schulman, and Wojciech Zaremba, among others. The OpenAI platform provides organizations with application programming interfaces (APIs) they can then use to develop new generative AI-powered solutions that leverage all the advanced AI work and infrastructure OpenAI has already developed
Matplotlib	Matplotlib is a cross-platform, data visualization and graphical plotting library (histograms, scatter plots, bar charts, etc) for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications. A Python matplotlib script is structured so that a few lines of code are all that is required in most instances to generate a visual data plot. Matplotlib is a powerful and very popular data visualization library in Python. Matplotlib is very flexible and customizable for creating plots

Streamlit	Streamlit is an open-source Python framework for data scientists and AI/ML engineers to deliver dynamic data apps with only a few lines of code. Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. Streamlit allows to create a stunning-looking application with only a few lines of code. It is a Python-based library specifically designed for machine learning engineers. Streamlit is a powerful and user-friendly open-source Python library that makes it easier to build interactive web applications for machine learning and data science. With Streamlit, developers and data scientists can create engaging, informative, and visually appealing apps with just a few lines of code.
Pandas AI	Pandas AI is the best generative AI Python library for data scientists and machine learning professionals. It is an extension of the Pandas library that improves standard data wrangling functions with a next-generation generative pre-trained AI. The utilization of the newest AI algorithms in Pandas AI provides users with an opportunity to carry out operations on large amounts of data using the least amount of time and error margin, thus making it an essential tool for data manipulation in the contemporary world. Hence, one of the most significant strengths of Pandas AI is its compatibility with OpenAI, which is qualified for natural language processing and machine learning. This integration enables the users to highly exploit the two libraries based on the strengths of each, seeking to provide an all-inclusive package in the use of data for user projects. Pandas AI is highly useful for building models for predictive analytics, data visualization, exploratory data analysis, and others irrespective of the variety of forms of analyses.

3.2 The Prototype Solution

Below diagram illustrates the process of data visualization using OpenAI



The below flowchart represents the Streamlit-based data visualization application's workflow. Each section of the workflow is clearly depicted with labeled icons and a modern design

The Python solution developed for OpenAI provides a user-friendly platform for data visualization and analysis using OpenAI's capabilities. Key features include:

1. **CSV Upload and Preview**: Users can upload CSV files for analysis.
2. **OpenAI-Powered Insights**: Allows users to query their data using natural language prompts.
3. **Custom Visualizations**: Generates interactive charts based on specific prompts.

4. **Dynamic Styling**: Includes dark-mode themes and animations for a modern look and feel.

Below are the breakdown of the features of the solution developed

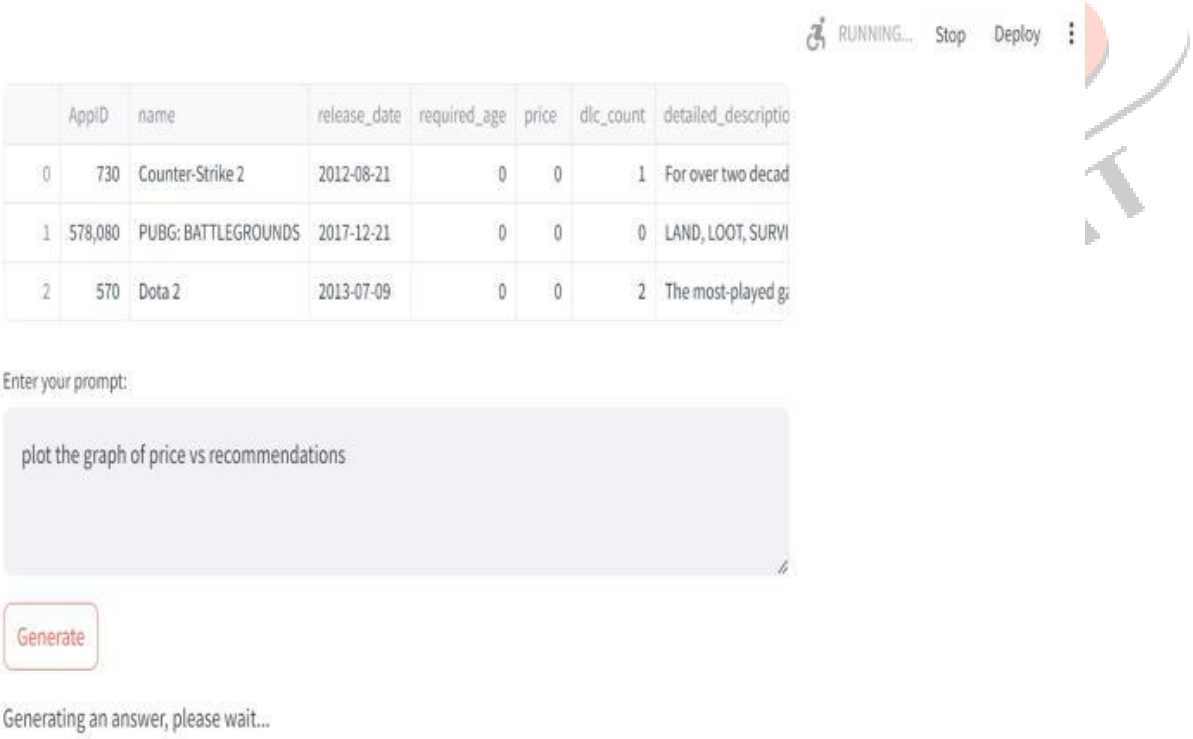
1. **File Upload and Preview** - Accepts a CSV file and displays the first three rows as a preview.
   Enhancement Idea: Add file validation to ensure the uploaded file has no structural issues.
   Show metadata (e.g., the number of rows and columns, missing values).

2. **OpenAI-Powered Prompts** - Uses OpenAI's capabilities to interpret user queries about the dataset. OpenAI was leveraged to translate user queries into visualizations. Example: When a user asked, "What are the top 5 games by sales?" OpenAI provided the necessary SQL query and instructions for rendering a bar chart using libraries like Matplotlib

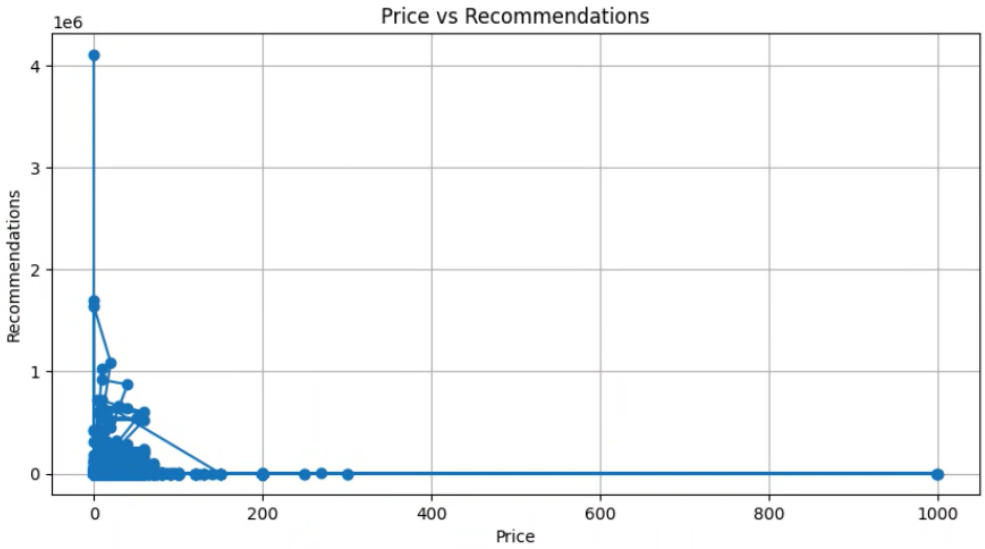
3. Visualization Handling - Creates a sample plot if columns such as price and recommendations exist.
   Enhancement Idea: Dynamically list available columns for users to choose from.
   Add support for more visualization types (e.g., scatter plots, heatmaps).
   Use libraries like **Altair** or **Plotly** for more interactive charts

3.3 Results

Output Visualization as below



Generating an answer, please wait...



3.4 Interpretation

This solution demonstrates a robust foundation for combining Streamlit and OpenAI for data analysis and visualization. By incorporating feedback and enhancements, the tool can be made more versatile, interactive, and user-friendly for both technical and non-technical users

3.5 Conclusions

- **Power of OpenAI in Visualization**: OpenAI's capabilities provide transformative tools for data visualization that can be used across a wide range of fields, from business analytics to scientific research.
- **User-Centric Design**: As OpenAI tools improve, it is clear that user experience and customization will play an important role in their success.
- **Data Accessibility**: These tools have significantly improved the accessibility of complex data, allowing users without advanced technical skills to create insightful visualizations

3.6 Directions for Future Research

- **Exploring Unstructured Data**: Future research should focus on using OpenAI to visualize unstructured data such as text, audio, and images, and to generate meaningful insights from them.
- **AI Integration in Big Data**: Investigating how OpenAI can integrate with big data platforms to enhance visualization of large datasets.
- **Cross-Platform Visualization**: Exploring how OpenAI models can be applied to create visualizations across different platforms (e.g., mobile apps, desktop, VR).
- **Generative Visualizations**: Research into the use of generative AI techniques to produce new and innovative data visualization formats.

As OpenAI and similar models improve, there is potential for more advanced data interpretation, including multi-modal data inputs (e.g., combining text, images, and numerical data for analysis). The ability to process complex, multi-modal data could open up new avenues for richer insights and visualizations, further pushing the boundaries of automated data analysis.

Future versions of OpenAI can be expected to generate more interactive, real-time visualizations, allowing users to explore and modify visualizations dynamically based on evolving queries. This would enable a more exploratory and hands-on approach to data analysis, where users can continuously refine their queries to uncover deeper insights.

REFERENCES

[1]Amershi, S., Counsell, S., & Teytelman, A. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-13. Association for Computing Machinery. doi:10.1145/3290605.3300233

[2]Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.

Heer, J., & Shadbolt, N. (2017). Interactive Data Visualization: A Review of the Literature. *Computers & Graphics*, 73, 65-79. doi:10.1016/j.cag.2017.08.005

[3]Kirk, A. (2016). *Data Visualization: A Handbook for Data Driven Design*. Sage Publications.

Sacha, D., & et al. (2018). The Role of Interaction in Data Visualization. *Information Visualization*, 17(2), 115-129. doi:10.1177/1473871617740212

[4]Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press.

[5]Huang, Y., Xie, Z., & Wang, Y. (2020). User-Centered Design for Data Visualization: A Review. *IEEE Transactions on Visualization and Computer Graphics*, 26(9), 1-13. doi:10.1109/TVCG.2019.2935134

[6]Ware, C. (2013). *Information Visualization: Perception for Design* (3rd ed.). Morgan Kaufmann.

[7]Borkin, M. A., & et al. (2013). Beyond Color: A Taxonomy of Data Visualization User Interfaces. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2211-2220. doi:10.1109/TVCG.2013.203

[8]Chen, M., & et al. (2014). Data-Driven Visualization for Interactive Analysis. *Journal of Computational and Graphical Statistics*, 23(1), 111-126. doi:10.1080/10618600.2013.830866

[9]Jiang, T., Liu, E., Baig, T., & Li, Q. (2024). Enhancing decision-making in higher education: Exploring the integration of ChatGPT and data visualization tools in data analysis. *New Directions for Higher Education*, 2024, 15–29. https://doi.org/10.1002/he.20510

[10]Brown, T., Mann, B., Ryder, N., & et al. (2020). *Language Models are Few-Shot Learners*. In *Advances in Neural Information Processing Systems* (NeurIPS 2020).

[11]OpenAI. (2021). *DALL·E: Creating Images from Text*.

[12]OpenAI. (2021). *CLIP: Connecting Text and Images*.

[13]Miller, C. & et al. (2022). *Integrating OpenAI Models into Data Visualization Workflows*. In *Proceedings of the IEEE Conference on Visualization* (IEEE VIS 2022).

[14]Liu, H. & et al. (2022). *Data-Driven Visualization with OpenAI: A Comprehensive Framework*. *Journal of Data Science and Visualization*, 3(1), 45-58. doi:10.1016/j.jdsav.2022.01.004.

[15]OpenAI. (2022). *Exploring Data Visualizations Using OpenAI's API*.

[16]Hernandez, J., & et al. (2023). *AI-Assisted Data Visualization: Leveraging OpenAI for Insight Generation*. *Data Visualization Journal*, 8(1), 34-50. doi:10.1177/2047487323121001.