Question 1 (66 pts)

1. (4 pts) Import the Chicago neighborhood data that has been provided in D2L. Every row of data represents a unique Chicago neighborhood and contains information about its Name, the percentage of its Black population, the percentage of its Hispanic population, its Unemployment rate, and its median Income level. Upon a successful import you should see 5 variables and 76 observations.

A screenshot of a computer

Description automatically generated

2. (8 pts) Calculate summary statistics for all the numerical variables in the data set.

A close-up of numbers

Description automatically generated

3. (4 pts) Use the set.seed() command to set the starting value to 23. This ensures that any random sample generation starts at a pre-determined value (in this case from 23), causing R to generate exactly the same random sample every time your code runs. This is very useful for model building and ensures that the clusters are created in the same order every time the code runs.

A close up of a text

Description automatically generated

4. (10 pts) Clustering is an unsupervised machine learning technique that is used to find hidden patterns in the data. In this section we will utilize the Elbow Method to choose the appropriate number of clusters based on minimizing the within sum of squares. Utilizing all the numerical variables in the data set, calculate and plot the within sum of squares for cluster sizes 1 through 15. As we have done in class, your plot should have the number of clusters on the horizontal axis and the within sum of squares on the vertical axis.



5. (6 pts) According to the elbow graph, what is the appropriate cluster size to choose?

3 Clusters since adding more clusters doesn’t significantly reduce the within-cluster variance.

6. (8 pts) Estimate a clustering model using the K-Means algorithm by setting the number of clusters to the approriate cluster size that you identified above.



7. (6 pts) Tabulate the cluster centers for the estimated K-Means model.

A number of numbers and symbols

Description automatically generated with medium confidence

8. (10 pts) Create a plot of the clusters using ggplot, with Unemployment on the horizontal axis, and Income on the vertical axis. Make sure your plot differentiates between the clusters using different colors.



9. (10 pts) Looking at the clusters centers and the cluster plot, what can you say about the racial distribution of income and unemployment in the Chicago neighborhoods? Briefly explain

K-means clustering plot of Chicago neighborhoods based on **income** (y-axis) and **unemployment rate** (x-axis). The clusters are color-coded, with three distinct clusters represented by different colors: red, blue, and purple

* **Cluster 1 (Red)**: This cluster represents neighborhoods with **higher income** (above $100,000) and **lower unemployment rates**. These neighborhoods likely have more affluent populations with greater economic stability.
* **Cluster 2 (Blue)**: These neighborhoods have **moderate income levels** (around $60,000 to $90,000) and **moderate unemployment rates**. These areas could represent middle-income communities.
* **Cluster 3 (Purple)**: This cluster corresponds to neighborhoods with **lower income levels** (below $50,000) and **higher unemployment rates**. These areas may be economically disadvantaged, facing higher unemployment and lower wages.

The clustering analysis suggests a clear socioeconomic gradient across Chicago neighborhoods:

* The red cluster represents wealthier areas with low unemployment, likely corresponding to predominantly white or affluent populations.
* The blue cluster represents middle-income areas with moderate unemployment, which may have a more racially diverse population.
* The purple cluster represents economically disadvantaged areas with high unemployment and low income, potentially corresponding to neighborhoods with higher concentrations of minority populations.

Question 2 (34pts) Please watch the documentary ”In the Age of AI” that has been posted on D2L to answer the questions below.

1. (4 pts) What was the Sputnik moment that led China to decide to up their AI game?

when Google's AlphaGo, a deep-learning AI program, defeated the world champion of the ancient Chinese board game Go, Lee Sedol, in 2016. This event was seen as a critical moment that made the Chinese government realize the importance of catching up in artificial intelligence.

2. (4 pts) What move number made by AlphaGo surprised grandmaster Lee Sedol the most? What was special about that move?

The move that surprised grandmaster Lee Sedol the most was move 37 in game two. This move was special because it was a move that humans could not understand but ended up being brilliant, revealing a strategy that human players had not considered.

3. (4 pts) In the new world where data is the new oil which country is the new Saudi Arabia? Why?

In the new world where data is the new oil, China is often referred to as the new Saudi Arabia. This is because China has a vast amount of data available, significantly more than other countries, which is crucial for the development and effectiveness of artificial intelligence (AI).

4. (4 pts) According to Wu Wenhao, how quickly can Megvii recognize a human face on a mobile device?

According to Wu Wenhao, Megvii can recognize a human face on a mobile device in less than 100 milliseconds, which is approximately 0.1 second.

5. (6 pts) According to Kai-Fu Li what percentage of jobs are likely to be somewhat or extremely threatened by AI over the next 10-15 years? Are these mostly blue collar jobs?

According to the transcript, Kai-Fu Lee believes that about 50% of jobs will be somewhat or extremely threatened by AI in the next 15 years or so. The impact of AI could be broader and affect various types of jobs. There is no specific mention in the video stating mostly blue collar jobs will be affected.

6. (6 pts) Why did productivity and income growth decouple in the past couple of decades in the United States? Why is this a bad thing?

The decoupling of productivity and income growth in the United States over the past couple of decades is a complex issue with several contributing factors. Key reasons include the rapid advancement of automation and artificial intelligence, which has significantly increased productivity by allowing machines to perform tasks more efficiently and accurately than humans. However, this has also led to job displacement and reduced the need for human labor in certain sectors, without corresponding wage increases for many workers. Additionally, income inequality has played a critical role, as the benefits of productivity gains have largely been captured by top earners and corporate profits, rather than being distributed more broadly among the workforce.

This decoupling is considered a bad thing because it exacerbates income inequality, reduces consumer spending, leads to economic instability, and negatively impacts social welfare. Lower wages mean less disposable income for consumers, which can dampen economic growth and lead to reduced consumer spending. The concentration of wealth among a few individuals and corporations can lead to economic instability, as it reduces the resilience of the economy to shocks and downturns. Furthermore, the lack of wage growth can lead to increased poverty and reduced social welfare, as more people struggle to make ends meet. This highlights the need for policies aimed at addressing income inequality and ensuring that the benefits of technological advancements are shared more broadly among the workforce

7. (6 pts) What is surveillance capitalism? Do you think it is dangerous? Why/why not?

Surveillance capitalism refers to an economic system where personal data is collected and commodified by corporations, often without users' explicit consent or awareness. This data is then used to create detailed profiles of individuals, which are used for targeted advertising, predictive analytics, and other forms of profit-making.The concept of surveillance capitalism was popularized by Shoshana Zuboff, who argues that it represents a new form of capitalism that is fundamentally different from traditional forms of capitalism. In surveillance capitalism, the primary commodity is not goods or services, but rather personal data, which is harvested and sold to third parties.

1. **Efficiency**: Surveillance capitalism can lead to more efficient and personalized services, as companies can use data to tailor their offerings to individual needs.
2. **Innovation**: The use of personal data can drive innovation and economic growth, as companies can use data to develop new products and services.
3. **Consent**: Many argue that individuals can choose to opt out of data collection or use privacy-enhancing technologies to protect their data.

But problem is who is handling the data makes all the difference if it is in the wrong hands it can lead to disasters like

1. **Loss of privacy**: Surveillance capitalism erodes individuals' right to privacy, as their personal data is collected and used without their consent.
2. **Manipulation**: The use of personal data for targeted advertising and predictive analytics can be used to manipulate individuals' behavior and influence their decisions.
3. **Inequality**: Surveillance capitalism can exacerbate existing social inequalities, as those who have access to more data and better analytics tools can gain an unfair advantage over others.
4. **Democracy**: The concentration of power in the hands of a few large corporations can undermine democratic institutions and values.

individuals are often not aware of the extent to which their data is being collected and used, and that the opt-out mechanisms are often inadequate or difficult to use.

China's social credit system, which uses data to monitor and control individuals' behavior. The documentary also notes that the use of facial recognition technology in China has raised concerns about privacy and civil liberties.

Overall, while surveillance capitalism can bring some benefits, its dangers and risks should not be ignored. It is essential to have a nuanced understanding of the implications of surveillance capitalism and to develop regulations and safeguards to protect individuals' privacy and rights.