

Final

MJ Johns Adam Khan

5.1.1 1. Introduction (0.75-1 page)

- Research question or paper overview
- Background and motivation (make sure to include proper citations)
- Starting sample reference (Yannakakis 2005)
- Dataset description (source, size, key variables)
- Study objectives

5.1.2 2. Exploratory Data Analysis (1-1.5 pages)

- Summary statistics
- Data visualization (2-3 key plots)
- Data quality assessment (missing values, outliers, distributions)
- Preliminary insights that inform modeling choices

Track 1 (Paper Reproduction) should include:

- A combination of original EDA and useful pieces of the EDA presented in the paper. If you use/reproduce a plot from the original paper in your report, give credit to the authors in the source.

5.1.3 3. Methods (1-1.5 pages)

Track 1 (Paper Reproduction) should include:

- Short explanation of the methods used in the original paper
- Proposed extension or alternative model (following the next section details)

Both tracks should include:

- **Statistical models used** with mathematical notation where appropriate
 - For regression: specify model equation, link function (if applicable), assumptions
 - For PCA: explain dimensionality reduction approach
 - For clustering: describe method and distance metric
 - For regularization: specify penalty type and selection procedure
- **Why these methods?** Connect to research questions and data structure
- Software and key R packages used (don't forget to cite the R packages)

5.1.4 4. Results (2-2.5 pages)

Track 1 (Paper Reproduction) should include:

- Summary of reproduction success/challenges
- What did the original authors do well? What could be improved?
- Summary of results with proposed model following the next section details.

Both tracks should include:

- Model fitting and diagnostics
 - Assumption checking (residual plots, normality tests, etc.)
 - Model comparison (if applicable)
 - Goodness-of-fit measures
- Parameter interpretation with confidence intervals where appropriate
- Key findings presented with visualizations

5.1.5 5. Discussion & Conclusion (0.75-1 page)

Both tracks should include:

- Answers to research questions with supporting evidence
- Practical implications or insights
- Limitations and assumptions
- Future directions

Track 1 (Reproduction) should include:

- A summary of what could be improved in the original paper and how your new analysis does exactly that.

5.1.6 6. References

- Not included in page count
- Use consistent citation style (APA, Chicago, or similar)
- Include at least 2-3 references beyond the dataset source

Yannakakis, Georgios N. 2005. "AI in Computer Games: Generating Interesting Interactive Opponents by the Use of Evolutionary Computation," December. <https://era.ed.ac.uk/handle/1842/879>.