

# Course Project Part 1

## Section A: Introduction to text embeddings on S&P 500 news

Criteria	Ratings	Pts
<b>Data Retrieval &amp; Structuring</b> a) Correctly retrieves S&P 500 tickers b) Collects news data and structures it into a DataFrame with required columns (TICKER, TITLE, SUMMARY, etc.)	<b>15 pts – Excellent</b> All data is accurately retrieved, well-structured, and fully cleaned in a professional DataFrame. <b>12 pts – Competent</b> Data structure is functional, though some cleaning or organization may be lacking. <b>9 pts – Basic</b> Partial data retrieval or significant issues in structure or completeness. <b>0 pts – Insufficient</b> Data not retrieved or structured correctly.	15 pts
<b>Embeddings &amp; Clustering</b> a) Embeddings generated with SentenceTransformer b) K-Means clustering applied c) Silhouette score used to determine best k d) Cluster labels added to DataFrame	<b>15 pts – Excellent</b> Embeddings are properly generated, clustering is applied with optimal k selection and integrated results. <b>12 pts – Competent</b> Embeddings and clustering are implemented, with some analysis gaps or minor flaws. <b>9 pts – Basic</b> Clustering or silhouette scoring is missing or incomplete. <b>0 pts – Insufficient</b> Embeddings and clustering are not implemented correctly.	15 pts
<b>Visualization</b> a) PCA used to reduce dimensions b) Clusters visualized in 2D with clear labeling and organization	<b>10 pts – Excellent</b> Visualization is clear, accurate, and well-labeled with meaningful insights. <b>8 pts – Competent</b> Visualization is mostly correct, but lacks clarity or has some missing labels. <b>6 pts – Basic</b> Visualization is present but unclear or not well executed. <b>0 pts – Insufficient</b> No visualization provided.	10 pts

<p><b>Reflection &amp; Analysis</b></p> <p>a) All technical, interpretive, and critical thinking questions are answered</p> <p>b) Responses demonstrate depth, understanding, and thoughtful analysis</p>	<p><b>10 pts – Excellent</b> Answers are complete, insightful, and reflect strong understanding of NLP and financial implications.</p> <p><b>8 pts – Competent</b> Most questions answered clearly, though depth or nuance is limited.</p> <p><b>6 pts – Basic</b> Superficial answers or incomplete reflections.</p> <p><b>0 pts – Insufficient</b> Reflection questions are missing or not meaningfully addressed.</p>	<p>10 pts</p>
---	--	---------------