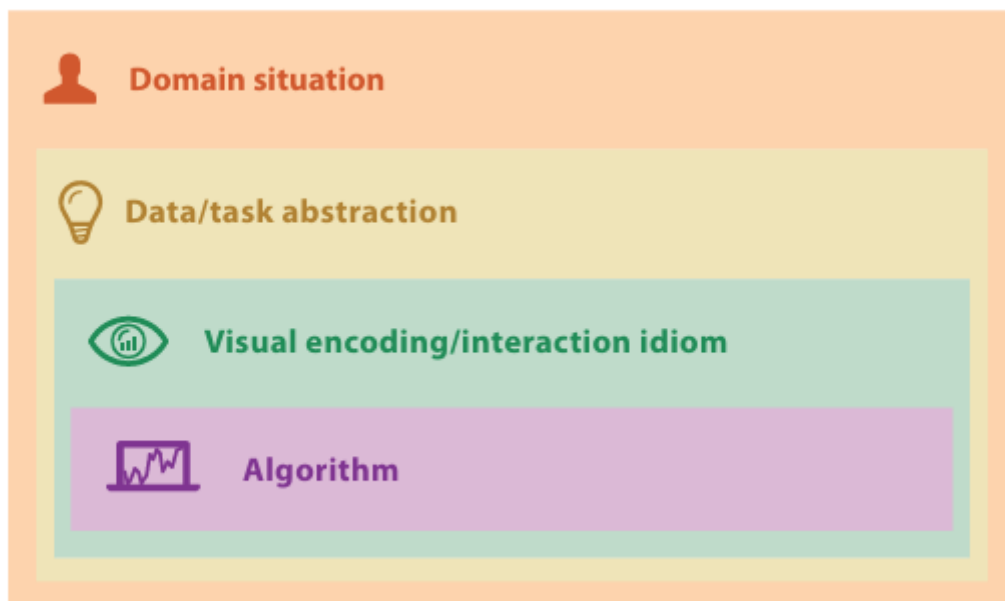# Instructions

Video Response Instructions

team name, team member names

Nested Model Summary:

- Domain Situation:

  Interactive Clustering of Readers' Behavior for News Articles

- Task Abstraction:

    - Clustering different types of readers based on their reading behavior

    - Analyzing article preferences of different reader clusters

    - Digging insights of relationship between reader behavior and article preferences

- Visual Encoding:

    - Core Visualization: Polar Plot, A clock-like plot with 24 hours on the theta axis indicating the time accessing articles, and the radius axis, indicating the difference of characteristics of readers, showing the distribution of different reader clusters.

    - Supporting Visualization: Typical Charts, such as bar charts, scatter plots, and pie plot, showing the reading behavior, i.e., the duration, the article reading coverage, and bookmark/favorite rate of different reader clusters against different article categories.

    - Algorithm:

        - Clustering: K-means, to cluster readers based on their reading behavior, using SKlearn

- Dimensionality Reduction: PCA, to help visualization of the difference of reading behavior characteristics, which is high-dimensional, using SKlearn

A reminder (or update) of your tool's task list:

```
-  Task Abstraction: (same as above)

   - Clustering different types of readers based on their reading behavior

   - Analyzing article preferences of different reader clusters

   - Digging insights of relationship between reader behavior and article
preferences
```

- Ademonstration of your tool's functionalities, with voiceover description.

A clear characterization of the visual encoding, interactions, and view composition that your visualization(s) employ, with demonstration and voiceover description that clearly ties these decisions back to your project's task, domain, and user characterization (5 points)

- Core visualization is designed as a overview of the users distribution, also a entry point for tool users to select a specific user cluster for further analysis.

- Supporting visualizations are designed to provide detailed information:

  - A hist chart indicating the counts of time readers access articles, allowing a distribution comparison among weekday commuter readers, weekend readers, and night readers.

  - Pie charts showing the distribution of article categories, allowing a comparison of the reading preferences of different reader clusters.

  - Scatter plots showing the reading depth distribution of different reader clusters:

    - x-axis: the average reading durationtime of articles
    - y-axis: the average reading coverage of articles
    - dot color: the category of articles

A reflection on how well your interaction and encoding decisions support your tasks, with discussion on what you could refine to better support them (1.5 points)

Core Visualization Improvements

- Axis Optimization:

  - Theta axis: better define the time of every user, the current average time is not the best option to represent the access time characteristic of users.

  - Radius axis:

    - Better dimensionality reduction, mainly optimizing the input multi-dimensional data;

- Give selections of single characteristics as the radius axis.

- Cluster Number Selection: Provide a selection of the number of clusters, allowing users to choose the best clustering result.

Supporting Visualization Improvements

```
Add more meaningful visualizations, and interactions.
```