IS 669 Big Data and Information Systems
Fall 2022

Midterm Exam

DATA EXPLORATION AND ANALYSIS

Professor HG Locklear
hlocklear@pace.edu

Exam Instructions

- This is your midterm exam, and it is designed to evaluate your ability to perform simple data operations in Python.
- Utilize the provided data file to complete all tasks.
- Complete all task within a single Jupyter Notebook and organize the notebook in Task order. (Data Cleaning, Data Visualization, Data Analysis, Data Modelling, and Data Sampling)
- Provide HTML markup in your notebook for each Task so that it is easily understood and organize any helper functions, for that Task, so that they are easily identified.
- The evaluation of your solution will strongly consider the **readability** of your notebook as well as the **correctness** of your code.
- You are allowed to **import only** the **dataclass**, **random**, and **matplotlib** modules/libraries.
- You **may not** use any of the built-in Python functions for statistics.

The Data Problem

- All Big Data problems involve the use of data sets to provide the raw values to build information about a particular entity.
- This information is often then used to model the entity so that a deeper analysis can be conducted into its behavior.
- Our problem involves the analysis of StarCraft2 Target data which will be used to assist AI developers in creating 'intelligent' game bots that can play against humans in World-wide competitions.
- You can learn more about StarCraft2 AI competition here **https://sc2ai.net/**

- **Our task is to utilize a text file containing corrupted StarCraft2 Target data and create a series of processes that will allow our AI bot engineers, here at Pace University, to better develop their bots.**

StarCraft 2 Target Data

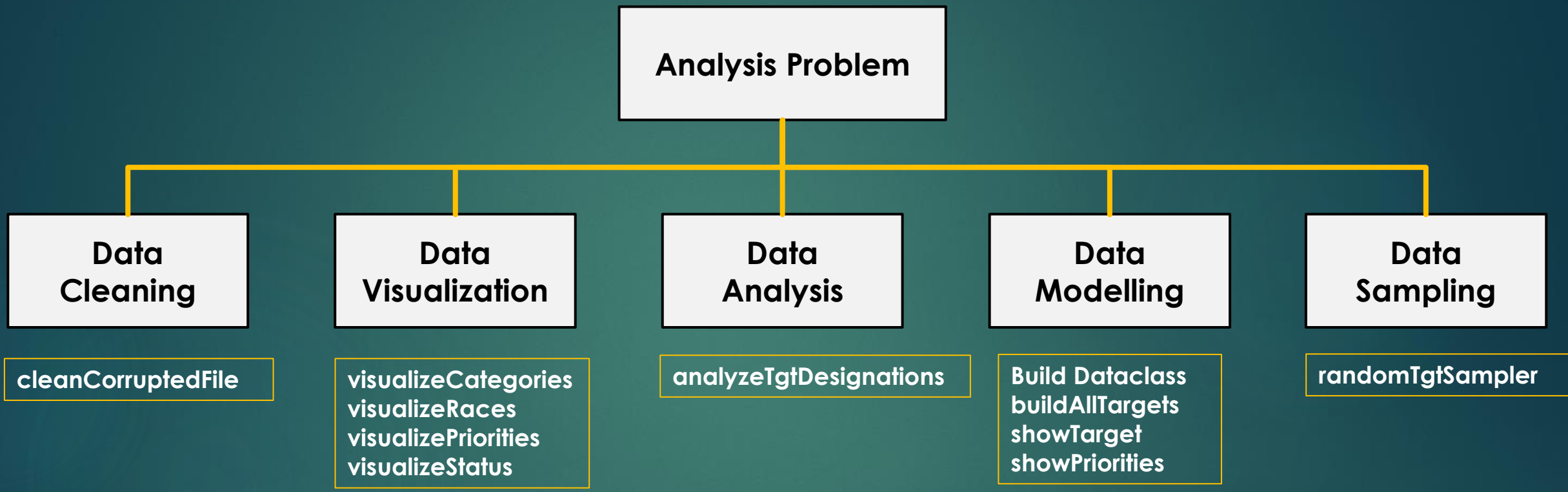
- Each line in the data file represents an **individual StarCraft 2 target**.
- Each space separated value corresponds to a **target attribute.**

```
AB-75-4858 Protoss High Rectangle 94 62 3 Tactical True System Concentrated Mobile Protected 765 696 Airborne 140 99 Active 272
```

Target Attributes (Left to Right)

1. Target Designation
2. Target Race
3. Target Priority
4. Target Shape
5. Target Length
6. Target Width
7. Number of Target Components
8. Target Category
9. Target Surveillance
10. Target Type
11. Target Density
12. Target Mobility
13. Target Protection
14. Target X Coordinate
15. Target Y Coordinate
16. Target Observer
17. Target Discovered Game Time
18. Target Health
19. Target Status
20. Target Point Value

Problem Overview



Data Cleaning

- Create the function **cleanCorruptedFile** that accepts the corrupted StarCraft2 Target data file and the name of a file to write the clean data to.
- The function removes the corruption in the file and produces a new file.

SC2TGT_ListCorrupted

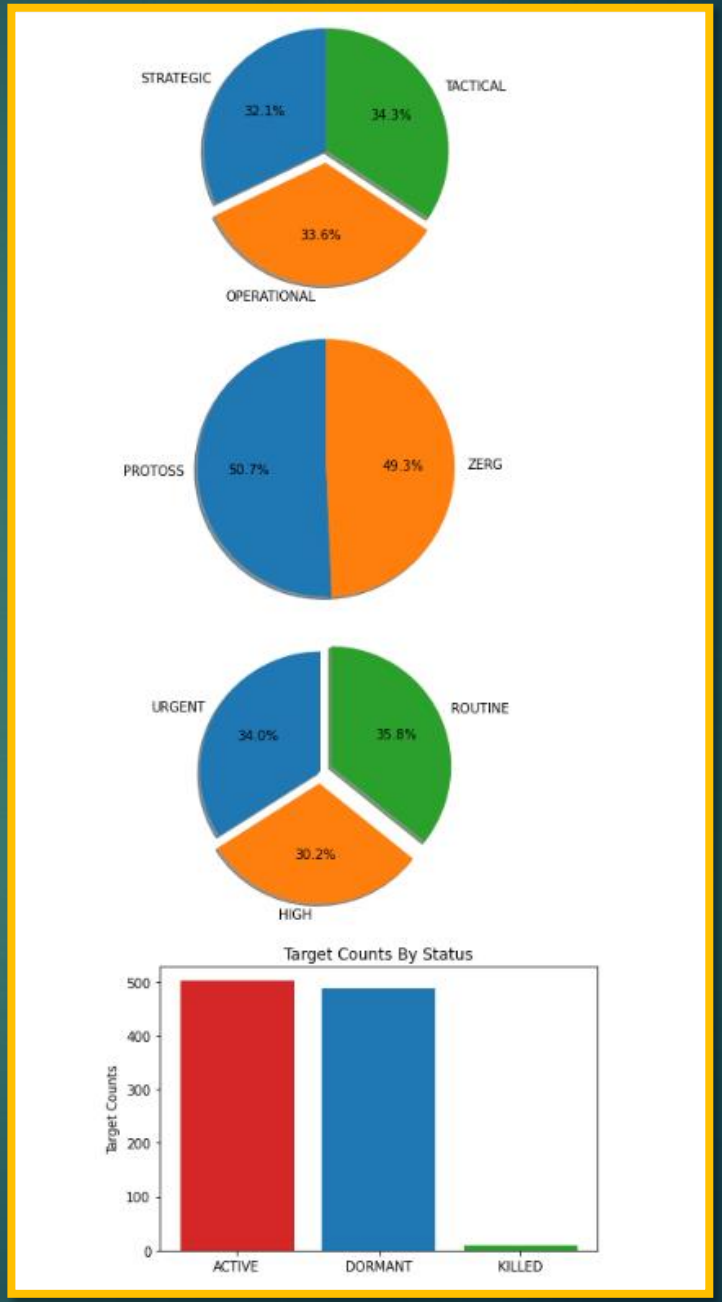
```
$%*@AD-25-2995 Zerg Urgent Square 62 121 7 Strategic False System Concentrated Mobile Shielded 213 381 Ground 30 29 Active 271##**
$%*@AD-100-3071 Zerg High Rectangle 119 133 9 Operational False Unit Dispersed Mobile Fortified 102 282 Airborne 53 54 Dormant 313##**
$%*@AA-34-9936 Zerg Urgent Polygon 78 143 3 Tactical False System Concentrated Mobile Protected 249 477 Satellite 80 85 Dormant 361##**
$%*@AB-57-5283 Zerg Routine Oval 262 91 1 Operational False Fortress Concentrated Static Exposed 269 289 Ground 123 69 Active 887##**
$%*@AF-57-8507 Protoss Urgent Square 70 70 11 Tactical True System Concentrated Mobile Shielded 751 868 Satellite 77 85 Dormant 112##**
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
**##963 evitcA 54 361 enrobriA 006 466 dedleihS eliboM detartnecnoC elcatsbO eslaF lacitcaT 6 98 27 elgnatceR hgiH ssotorP 0535-85-DA@*%$
$%*@AB-25-5001 Protoss Routine Rectangle 65 92 4 Tactical False Resource Dispersed Mobile Protected 930 653 Satellite 170 73 Dormant 225##**
```

SC2TGT_ListClean

```
AD-25-2995 Zerg Urgent Square 62 121 7 Strategic False System Concentrated Mobile Shielded 213 381 Ground 30 29 Active 271
AD-100-3071 Zerg High Rectangle 119 133 9 Operational False Unit Dispersed Mobile Fortified 102 282 Airborne 53 54 Dormant 313
AA-34-9936 Zerg Urgent Polygon 78 143 3 Tactical False System Concentrated Mobile Protected 249 477 Satellite 80 85 Dormant 361
AB-57-5283 Zerg Routine Oval 262 91 1 Operational False Fortress Concentrated Static Exposed 269 289 Ground 123 69 Active 887
AF-57-8507 Protoss Urgent Square 70 70 11 Tactical True System Concentrated Mobile Shielded 751 868 Satellite 77 85 Dormant 112
AD-58-5350 Protoss High Rectangle 72 89 6 Tactical False Obstacle Concentrated Mobile Shielded 664 600 Airborne 163 45 Active 369
AB-25-5001 Protoss Routine Rectangle 65 92 4 Tactical False Resource Dispersed Mobile Protected 930 653 Satellite 170 73 Dormant 225
```

Data Visualization

- Create the following functions
- **VisualizeCategories** which accepts a StarCraft2 Target List data file and produces a **pie chart** as shown.
- **VisualizeRaces** which accepts a StarCraft2 Target List data file and produces a **pie chart** as shown.
- **VisualizePriorities** which accepts a StarCraft2 Target List data file and produces a **pie chart** as shown.
- **VisualizeStatus** which accepts a StarCraft2 Target List data file and produces a **histogram** as shown.



Data Analysis

- Create the function **analyzeTgtDesignations** that accepts a StarCraft2 Target data file and provides (prints to shell) an analysis of the data as shown.

```
TOTAL TARGETS: 1000

TARGET DESIGNATION COUNTS
  AA: 167 AB: 164 AC: 149 AD: 173 AE: 149 AF: 198

TARGET TYPE COUNTS AA DESIGNATED TARGETS
  Resources: 22 Depot: 20 Barracks: 28 System: 25 Unit: 23 Obstacle: 27 Fortress: 22
TARGET TYPE COUNTS AB DESIGNATED TARGETS
  Resources: 24 Depot: 18 Barracks: 27 System: 22 Unit: 32 Obstacle: 19 Fortress: 22
TARGET TYPE COUNTS AC DESIGNATED TARGETS
  Resources: 22 Depot: 23 Barracks: 19 System: 24 Unit: 8 Obstacle: 22 Fortress: 31
TARGET TYPE COUNTS AD DESIGNATED TARGETS
  Resources: 17 Depot: 29 Barracks: 19 System: 31 Unit: 28 Obstacle: 25 Fortress: 24
TARGET TYPE COUNTS AE DESIGNATED TARGETS
  Resources: 23 Depot: 17 Barracks: 22 System: 22 Unit: 25 Obstacle: 22 Fortress: 18
TARGET TYPE COUNTS AF DESIGNATED TARGETS
  Resources: 31 Depot: 24 Barracks: 34 System: 24 Unit: 36 Obstacle: 27 Fortress: 22


TARGET PRIORITY COUNTS AA DESIGNATED TARGETS
  Urgent: 55 High: 51 Routine: 61
TARGET PRIORITY COUNTS AB DESIGNATED TARGETS
  Urgent: 56 High: 57 Routine: 51
TARGET PRIORITY COUNTS AC DESIGNATED TARGETS
  Urgent: 50 High: 44 Routine: 55
TARGET PRIORITY COUNTS AD DESIGNATED TARGETS
  Urgent: 61 High: 48 Routine: 64
TARGET PRIORITY COUNTS AE DESIGNATED TARGETS
  Urgent: 53 High: 46 Routine: 50
TARGET PRIORITY COUNTS AF DESIGNATED TARGETS
  Urgent: 65 High: 56 Routine: 77

ZERG TARGETS:
  AA: 86 AB: 75 AC: 84 AD: 80 AE: 66 AF: 102

PROTOSS TARGETS:
  AA: 81 AB: 89 AC: 65 AD: 93 AE: 83 AF: 96
```

Data Modelling

- Create the dataclass **SC2Target**

SC2Target	
designation: str race: str priority: str shape: str length: int width: int components: int status: str; pvalue: int health: int	mobility: str; protection: str density: str discovered: int discoverdBy: str coordinate_x: int coordinate_y: int; ttype: str surveillance: bool; category: str
display(): void	

```
IDENTIFICATION
  TGT ID: AD-25-2995 Race: Zerg Priority: Urgent
CONFIGURATION
  Shape: Square Length: 62 Width: 121 Aim Points: 7
CATEGORY
  Category: Strategic Surveillance: True Type: System
TARGETABILITY
  Density: Concentrated Mobility: Mobile Protection: Shielded
LOCATION
  X: 213 Y: 381
OBSERVER
  Located By: Ground At Time: 30
VALUE
  Health: 29 Status: Active Value: 271
```

display method prints to the shell the SC2Target's attributes in the format shown

Data Modelling

- Create the function **buildAllTargets** that accepts a StarCraft2 Target data file and creates an SC2Target object for each target in the file and returns all the SC2Target objects in a single list.

- Create the function **showTarget** that accepts a specific SC2Target designation and a list of all SC2Target objects and prints to the shell all the information about the specified target in the format shown on slide 9. If the specified target designation does not exist in the list, prints '**Target Does Not Exist**' to the shell.

Data Modelling

- Create the function **showTgtPriorities** that accepts a specific SC2Target priority value and a list of all SC2Target objects and prints to the shell the target counts for that priority in the format shown below.

```
Zerg Priority Routine
        Target Count is 173

Protoss Priority Routine
        Target Count is 185
```

Data Sampling

- Create the function **randomTgtSampler** that accepts a list of SC2Target objects and an integer and returns an analysis of the specified number of random SC2Target objects in the format shown

```
Sample Size: 10
RACE:   Zerg: 4 Protoss: 6

Zerg Total Target Point Value: 2401
Zerg Minimum Target Point Value: 453
Zerg Maximum Target Point Value: 761
Zerg Mean Target Point Value: 600.25
Zerg Target Point Value Range: 308

Protoss Total Target Point Value: 3494
Protoss Minimum Target Point Value: 128
Protoss Maximum Target Point Value: 933
Protoss Mean Target Point Value: 582.33
Protoss Target Point Value Range: 805
```