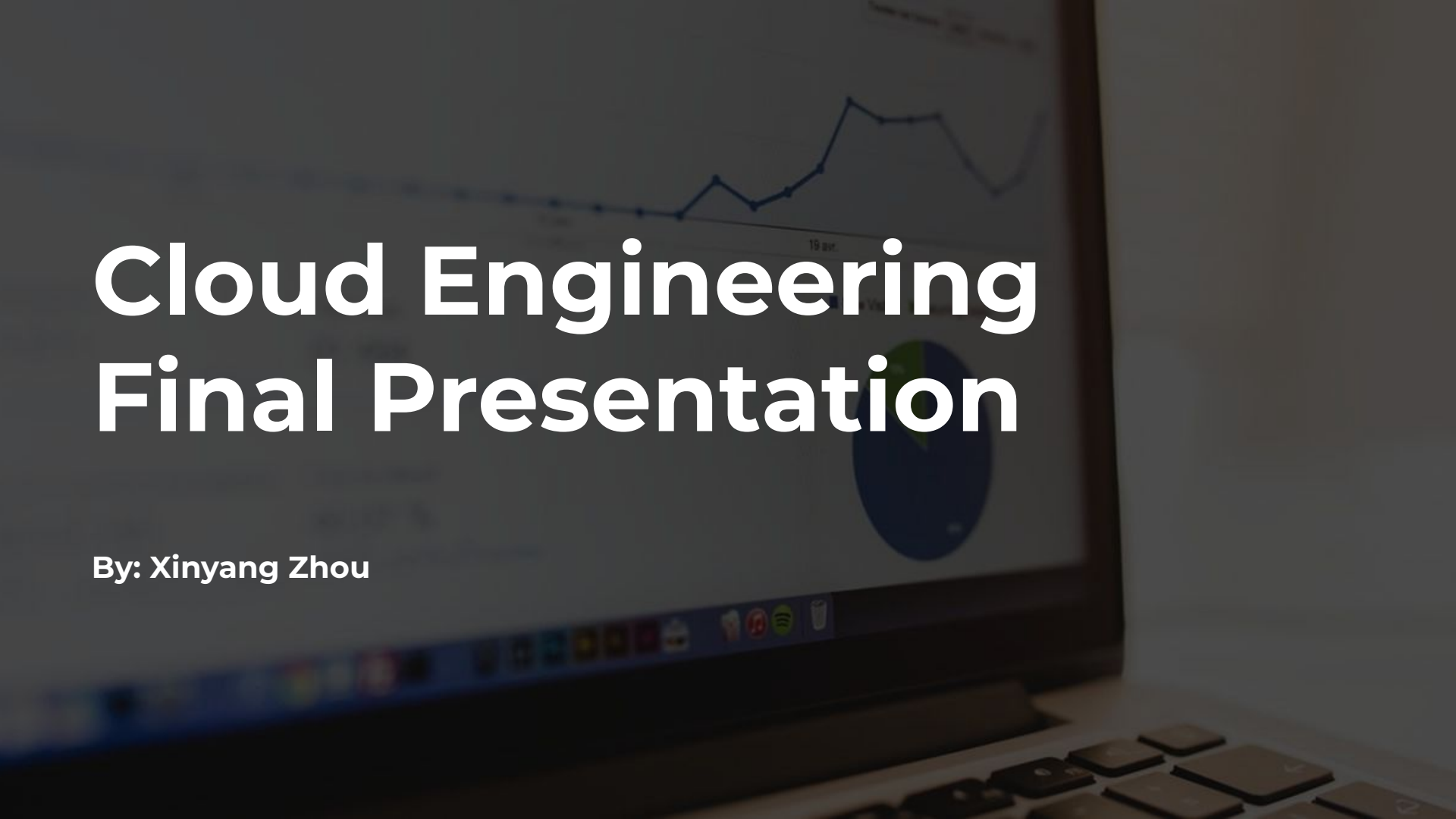


# Cloud Engineering Final Presentation

By: Xinyang Zhou

# Introduction: Project Goal

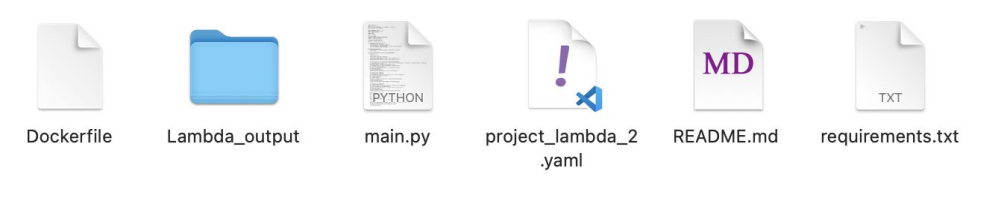
## Breast Cancer Prediction & Visualization via AWS

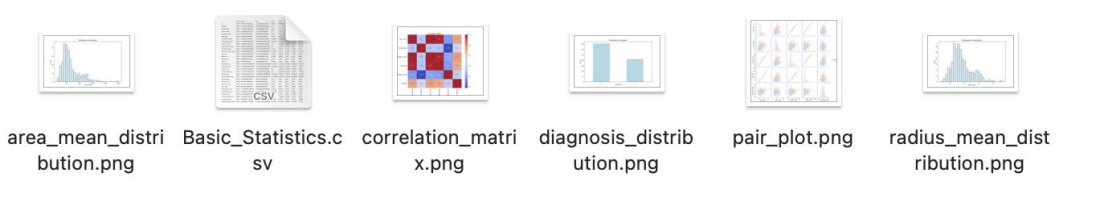
- A streamlined breast cancer detection, with real-time insights, on AWS, provided to related professionals.

- It's not just about spotting cancer; it's about making informed decisions and ultimately saving lives.

# Exploratory Data Analysis

- Using ECR and Lambda services on AWS.
- Utilize Dockerfile, YAML, AWS CLI, and other related tools covered in class.
- A Snapshot of the EDA section folder:



# Exploratory Data Analysis: Dataset

Basic_Statistics

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| id | 569.0 | 30371831.432337400 | 125020585.61222400 | 8670.0 | 869218.0 | 906024.0 | 8813129.0 | 911320502.0 |
| diagnosis | 569.0 | 0.37258347978910400 | 0.48391795640316900 | 0.0 | 0.0 | 0.0 | 1.0 | 1.0 |
| radius_mean | 569.0 | 14.127291739894600 | 3.5240488262120800 | 6.981 | 11.7 | 13.37 | 15.78 | 28.11 |
| texture_mean | 569.0 | 19.289648506151100 | 4.301035768166950 | 9.71 | 16.17 | 18.84 | 21.8 | 39.28 |
| perimeter_mean | 569.0 | 91.96903339191560 | 24.298981038754900 | 43.79 | 75.17 | 86.24 | 104.1 | 188.5 |
| area_mean | 569.0 | 654.8891036906860 | 351.914129181653 | 143.5 | 420.3 | 551.1 | 782.7 | 2501.0 |
| smoothness_mean | 569.0 | 0.0963602811950791 | 0.01406412813767360 | 0.05263 | 0.08637 | 0.09587 | 0.1053 | 0.1634 |
| compactness_mean | 569.0 | 0.10434098418277700 | 0.052812757932512200 | 0.01938 | 0.06492 | 0.09263 | 0.1304 | 0.3454 |
| concavity_mean | 569.0 | 0.0887993158172232 | 0.07971980870789350 | 0.0 | 0.02956 | 0.06154 | 0.1307 | 0.4268 |
| concave points_mean | 569.0 | 0.04891914586994730 | 0.038802844859153600 | 0.0 | 0.02031 | 0.0335 | 0.074 | 0.2012 |
| symmetry_mean | 569.0 | 0.18116186291739900 | 0.027414281336035700 | 0.106 | 0.1619 | 0.1792 | 0.1957 | 0.304 |
| fractal_dimension_mean | 569.0 | 0.06279760984182780 | 0.007060362795084460 | 0.04996 | 0.0577 | 0.06154 | 0.06612 | 0.09744 |
| radius_se | 569.0 | 0.40517205623901600 | 0.2773127329861040 | 0.1115 | 0.2324 | 0.3242 | 0.4789 | 2.873 |
| texture_se | 569.0 | 1.2168534270650300 | 0.5516483926172020 | 0.3602 | 0.8339 | 1.108 | 1.474 | 4.885 |
| perimeter_se | 569.0 | 2.8660592267135300 | 2.0218545540421100 | 0.757 | 1.606 | 2.287 | 3.357 | 21.98 |
| area_se | 569.0 | 40.337079086116 | 45.49100551613180 | 6.802 | 17.85 | 24.53 | 45.19 | 542.2 |
| smoothness_se | 569.0 | 0.007040978910369070 | 0.0030025179438390700 | 0.001713 | 0.005169 | 0.00638 | 0.008146 | 0.03113 |
| compactness_se | 569.0 | 0.025478138840070300 | 0.017908179325677400 | 0.002252 | 0.01308 | 0.02045 | 0.03245 | 0.1354 |
| concavity_se | 569.0 | 0.03189371634446400 | 0.03018606032298840 | 0.0 | 0.01509 | 0.02589 | 0.04205 | 0.396 |
| concave points_se | 569.0 | 0.011796137082601100 | 0.006170285174046870 | 0.0 | 0.007638 | 0.01093 | 0.01471 | 0.05279 |
| symmetry_se | 569.0 | 0.02054229876977150 | 0.008266371528798400 | 0.007882 | 0.01516 | 0.01873 | 0.02348 | 0.07895 |
| fractal_dimension_se | 569.0 | 0.003794903866432340 | 0.002646070967089200 | 0.0008948 | 0.002248 | 0.003187 | 0.004558 | 0.02984 |
| radius_worst | 569.0 | 16.269189806678400 | 4.833241580469320 | 7.93 | 13.01 | 14.97 | 18.79 | 36.04 |
| texture_worst | 569.0 | 25.677223198594000 | 6.146257623038320 | 12.02 | 21.08 | 25.41 | 29.72 | 49.54 |
| perimeter_worst | 569.0 | 107.26121265377900 | 33.602542269036400 | 50.41 | 84.11 | 97.66 | 125.4 | 251.2 |
| area_worst | 569.0 | 880.5831282952550 | 569.356992669949 | 185.2 | 515.3 | 686.5 | 1084.0 | 4254.0 |
| smoothness_worst | 569.0 | 0.13236859402460500 | 0.022832429404835500 | 0.07117 | 0.1166 | 0.1313 | 0.146 | 0.2226 |
| compactness_worst | 569.0 | 0.25426504393673100 | 0.157336488913742 | 0.02729 | 0.1472 | 0.2119 | 0.3391 | 1.058 |
| concavity_worst | 569.0 | 0.27218848330404200 | 0.2086242806081320 | 0.0 | 0.1145 | 0.2267 | 0.3829 | 1.252 |
| concave points_worst | 569.0 | 0.11460622319859400 | 0.06573234119594210 | 0.0 | 0.06493 | 0.09993 | 0.1614 | 0.291 |
| symmetry_worst | 569.0 | 0.2900755711775040 | 0.061867467537518700 | 0.1565 | 0.2504 | 0.2822 | 0.3179 | 0.6638 |
| fractal_dimension_worst | 569.0 | 0.0839458172231986 | 0.018061267348894000 | 0.05504 | 0.07146 | 0.08004 | 0.09208 | 0.2075 |

Important feature (target):

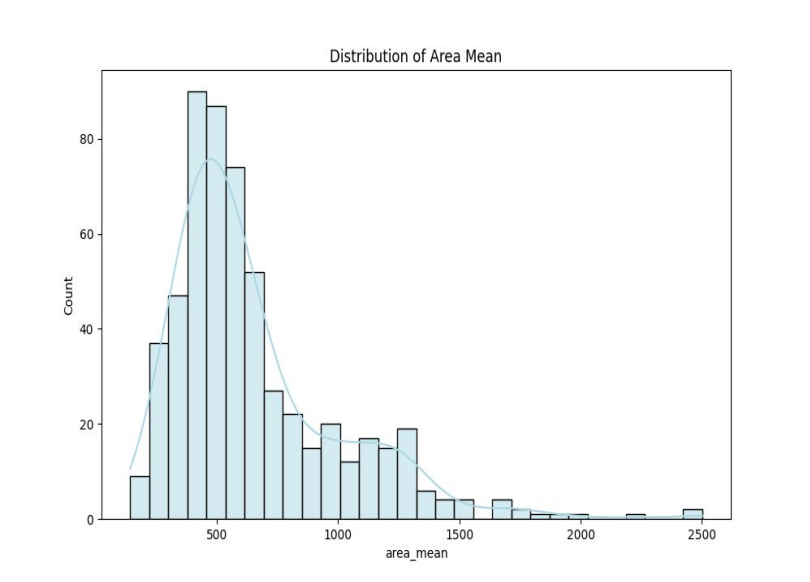
- “diagnosis”: 1-Cancer, 0-No Cancer

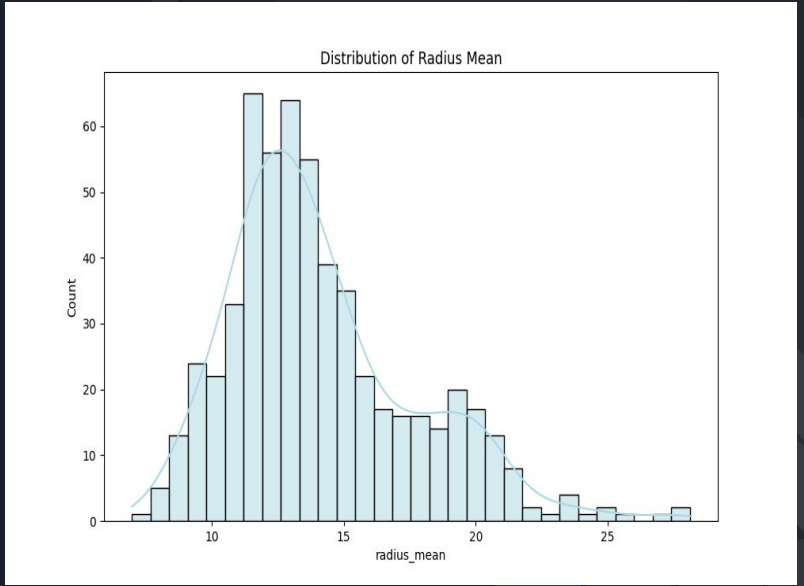
In the original dataset:

“B”-Benign, “M”-Malignant.

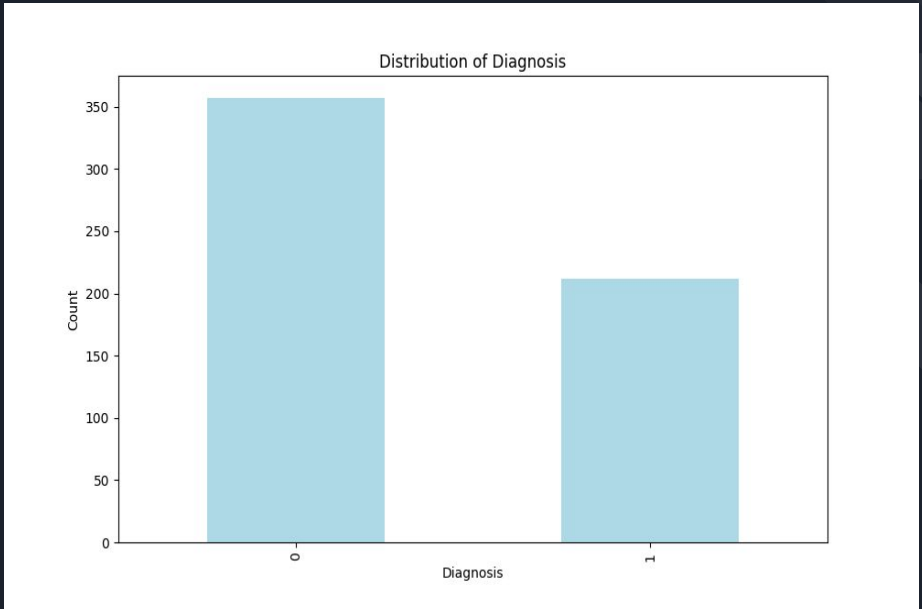
In the dataset, 30 columns are numerical, 1 column is string (target variable), 1 column is ID.

# Exploratory Data Analysis: Plots



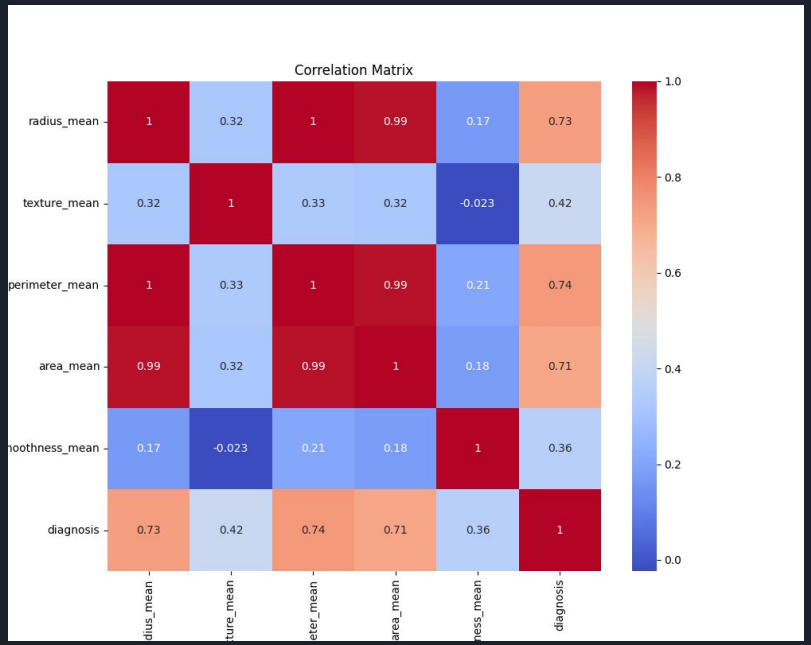



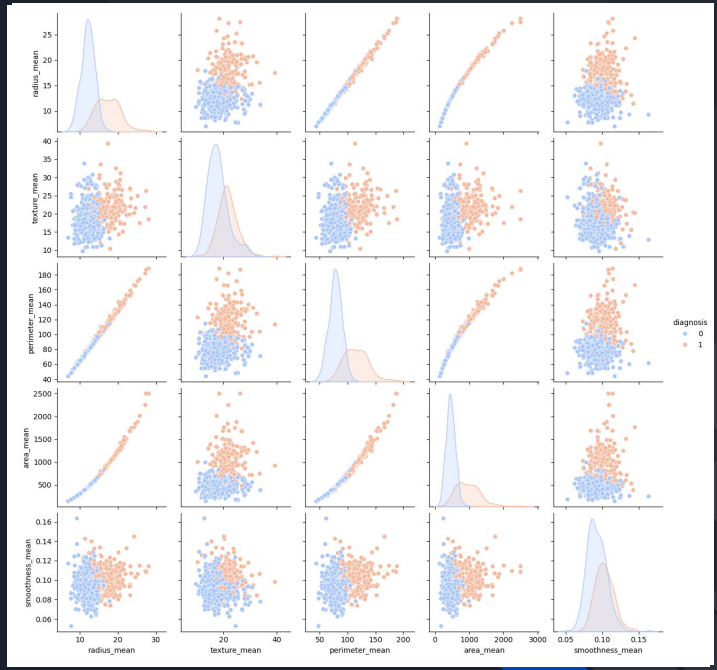
# Exploratory Data Analysis: Plots




There's more cases with no cancer in the dataset.
There are minimum concerns about class imbalance.

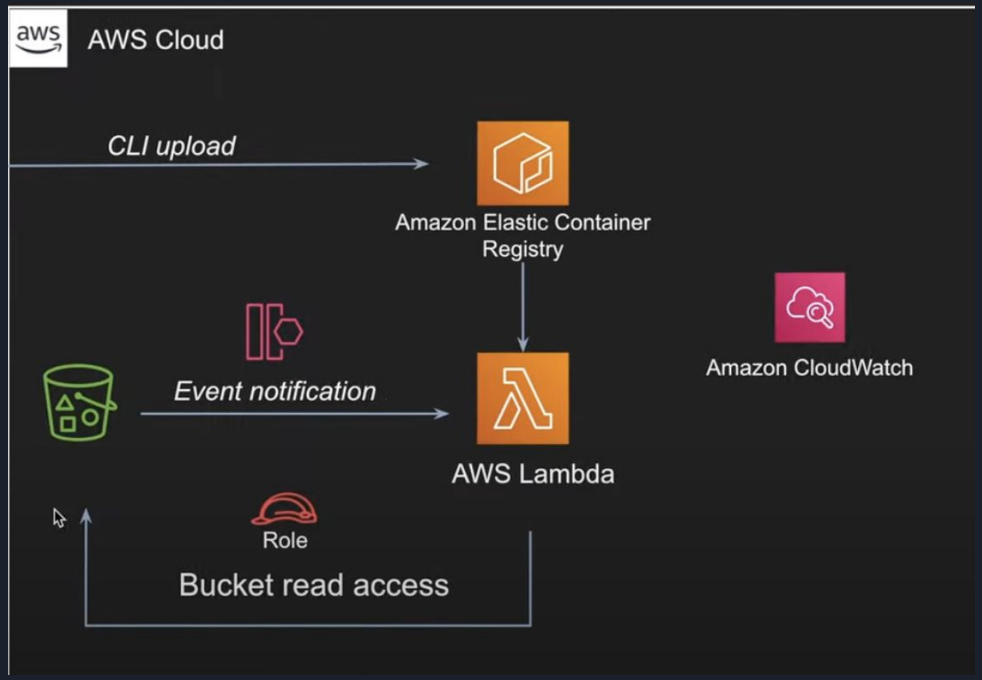
# Exploratory Data Analysis: relationships



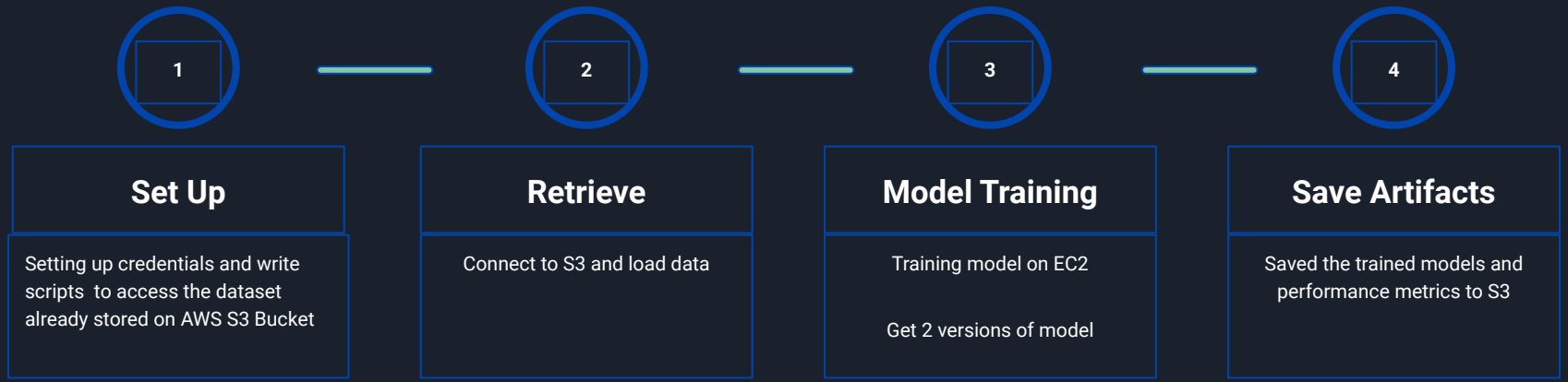




There are some strong correlations between variables.

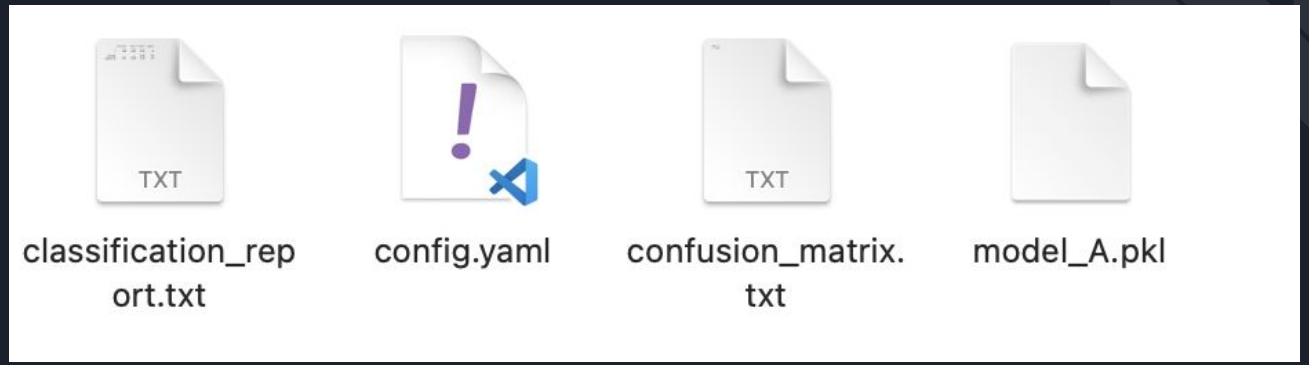
# Exploratory Data Analysis Section Diagram



# Next Step: Modeling on AWS



# Modeling Process Output

# Model Configurations

Model A: Random Forest Classifier:

(Smaller / Shallower tree)

```
n_estimators: 200
max_depth: 3
max_features: 3
```

Model B: Random Forest Classifier:

(Bigger / Deeper tree)

```
n_estimators: 500
max_depth: 10
max_features: 10
```

# Model Performance

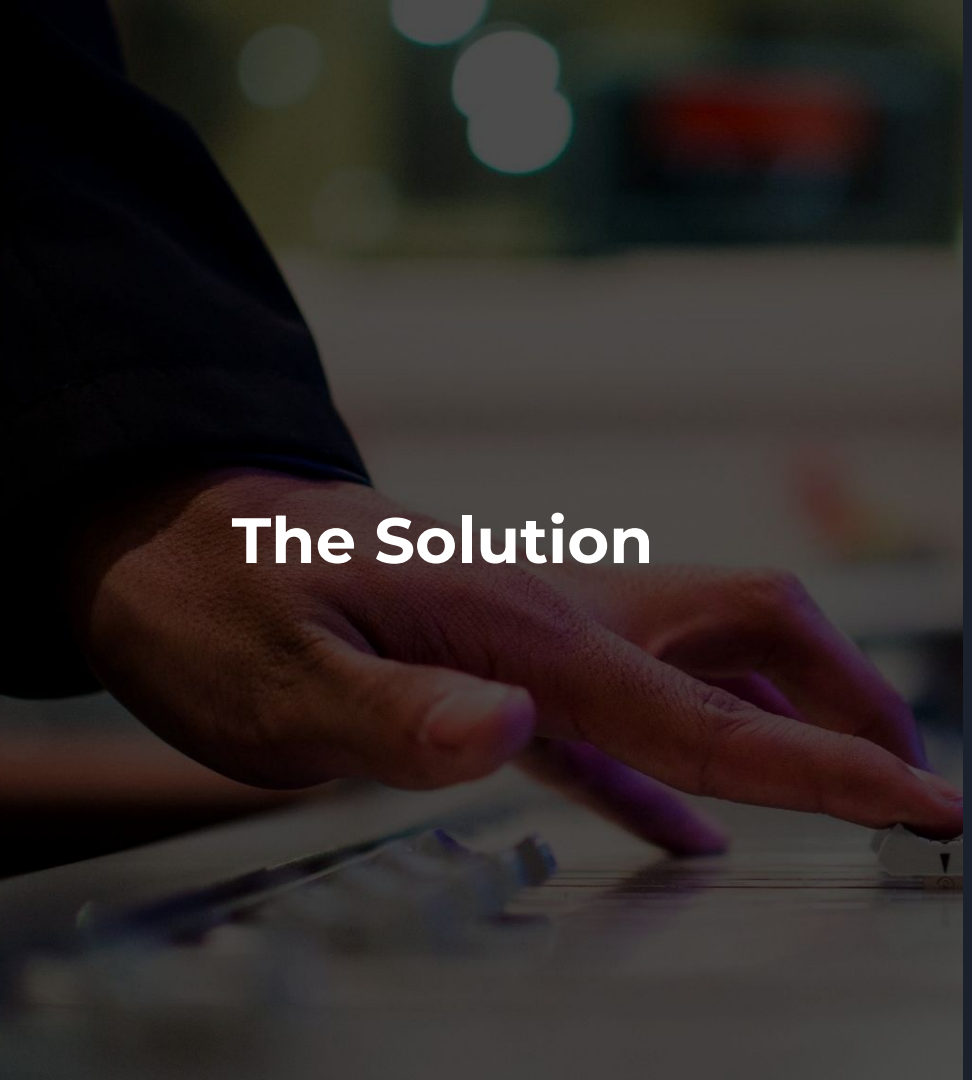
- **Model A:**

| | Precision | Recall | F1 Score | Support |
|---|---|---|---|---|
| B | 0.96 | 0.99 | 0.97 | 71 |
| M | 0.98 | 0.93 | 0.95 | 43 |
| Accuracy | | | 0.96 | 114 |
| Macro Avg | 0.97 | 0.96 | 0.96 | 114 |
| Weighted Avg | 0.97 | 0.96 | 0.96 | 114 |

# Model Performance

- **Model B:**

| | Precision | Recall | F1 Score | Support |
|---|---|---|---|---|
| B | 0.96 | 0.97 | 0.97 | 71 |
| M | 0.95 | 0.93 | 0.94 | 43 |
| Accuracy | | | 0.96 | 114 |
| Macro Avg | 0.96 | 0.95 | 0.95 | 114 |
| Weighted Avg | 0.96 | 0.96 | 0.96 | 114 |

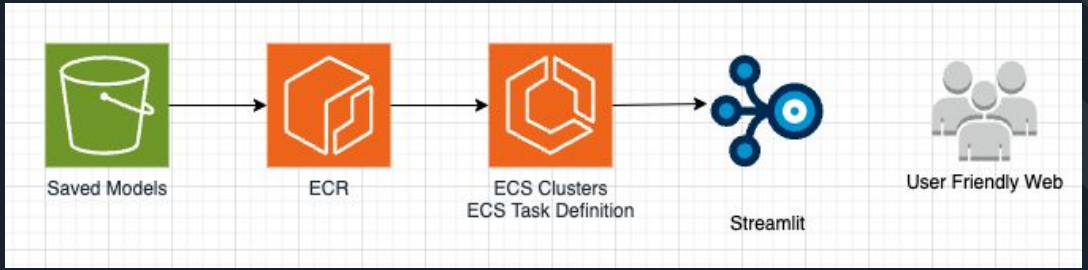


# The Solution

Using AWS services to build a user-friendly website that can be understood by everyone.

# Model Deployment

- Overall workflow:

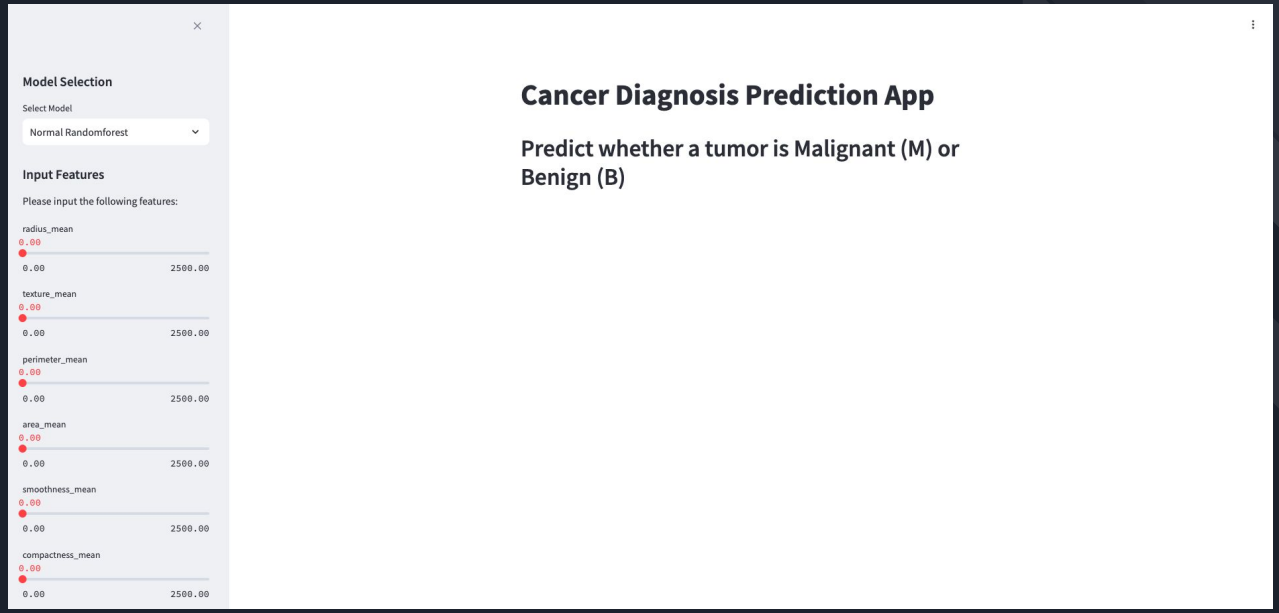


# Model Deployment

- Docker Image
- ECS Cluster
- ECS Task Definition
- Resources Configuration (service, environment variables, etc.)
- ECS Service
- Security Group Setup
- Grant Access to professionals and start making predictions instantly for your patients!

# Model Snapshot on the Web

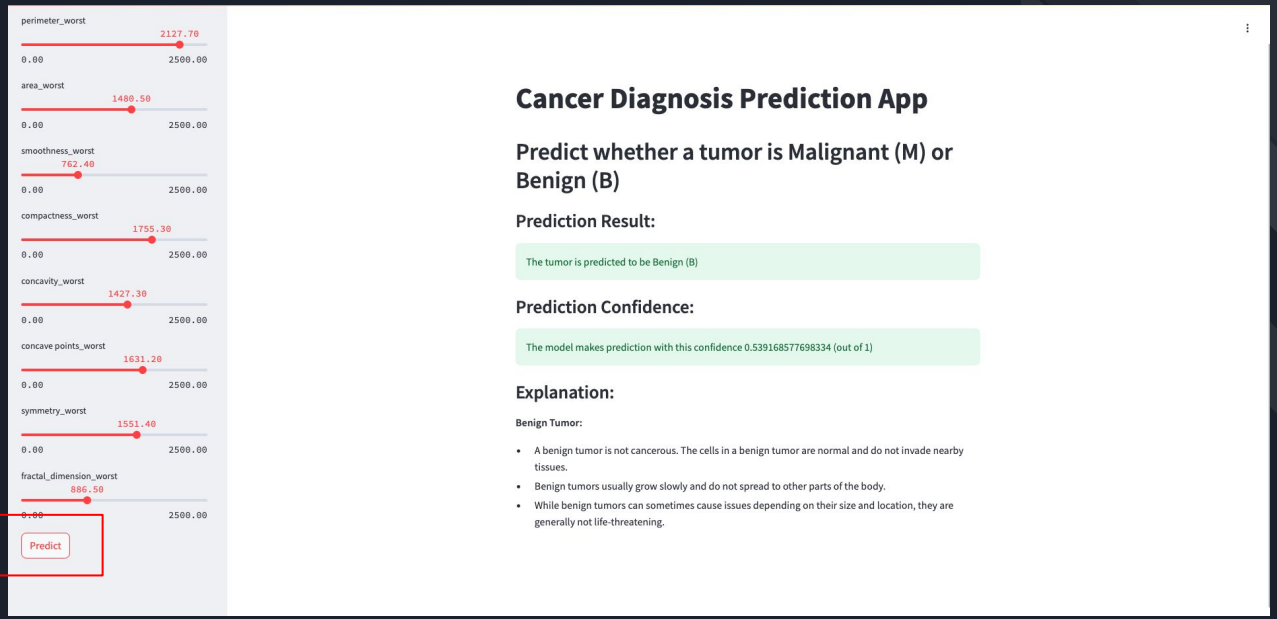
Landing Page:




Link: http://mlds-class-app-2094917527.us-east-2.elb.amazonaws.com/ (With Northwestern IP Address)

# Model Snapshot on the Web

After Clicking “Predict” Button:




Link: http://mlds-class-app-2094917527.us-east-2.elb.amazonaws.com/ (With Northwestern IP Address)


Thank you!