General Information

	Group Information
Group Number	4-13
Group Members	Bruyn Decker
	Gavin Ostheimer
	Kevin Paul
	Styles Weiler

NOTE: You can click the box with the url and an arrow in a box icon will appear at the bottom of the table to go to the site.	URL
Website URL	https://fagel-gamous.com/
WebsiteURL Backup (not required)	https://fagelgamous-test.us-east-1.elasticbeanstalk.com/
.ipynb URL (Supervised)	https://colab.research.google.com/drive/1CKasxp6vacqLUtk4fRKGKftEQiVwbTCG?usp=sharing
.ipynb URL (Unsupervised)	https://colab.research.google.com/drive/1reVemPlvUA_GMe64ZW7Ms7WgEEE620nU?usp=sharing

Username (email)	Password	Role	2FA/MFA Used (if any)
ta@is.com	FagelG@mous1	Admin	None (use for login)
MFA@is.com	FagelG@mous1	Admin	Yes (for demonstration. No login possible)

	URL	GitHub Branch
Repository URL	https://github.com/SinistralMars/Fag_el-Gamous.git	master

APRIL 14, 2023

FAG EL-GAMOUS PROJECT

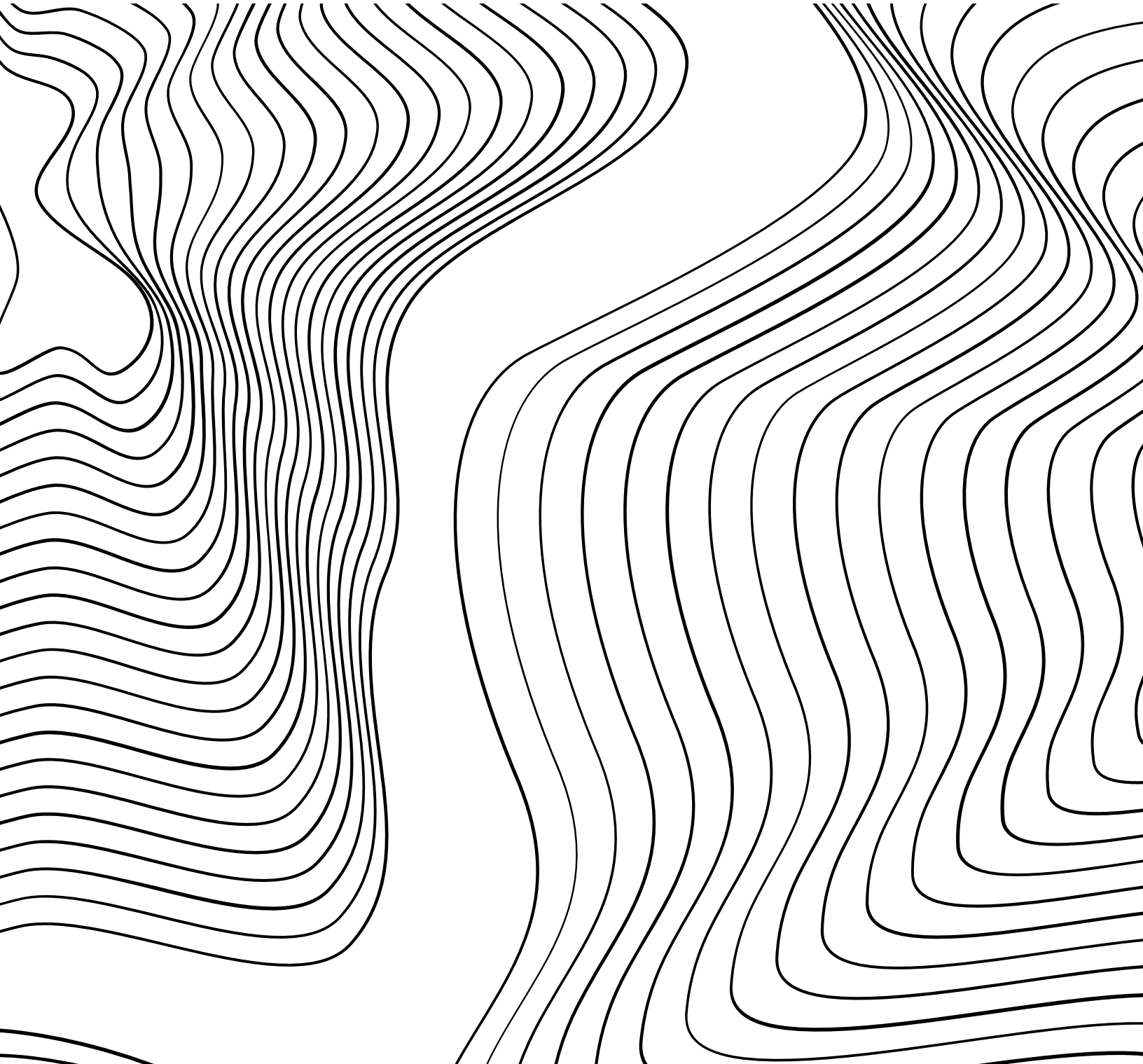


TABLE OF CONTENTS

01 Title

02 Table of Contents

03-05 Executive Summary

06 Website Tour

07-22 Documentation

07-10 IS404 Specifications

11 - 17 IS413 Specifications

18 - 20 IS414 Specifications

20 -22 IS455 Specifications

EXECUTIVE SUMMARY

The purpose of this project was to use analytic techniques to draw conclusions about burial digs in Fag el-Gamous, Egypt performed by BYU. The researchers requested us to complete an integrated exercise that achieved three overarching goals:

1. Further organize the data and develop a website where aspects of this information can be shared with researchers and interested members of the public around the world.
2. Allow the research team to view, filter, and query this data in many ways so they can answer a variety of academic research questions.
3. Use the data and advanced analytical/machine-learning skills to predict, explain, and visualize the wealth of data, allowing further insights.

We developed a .NET program in conjunction with advanced data analytics to accomplish these goals. First, we designed a secure site that allows the general to instantly access the wealth of data available through filtering and pagination. As for researchers, once authenticated they may add new findings to our dataset as well as edit and delete existing ones. Second, our Burial Summary page allows all users to view these mummies in a digestible format. We include the abilities to filter on up to 10 columns, which have been selected by the client. Third, our analysis focused on identifying patterns and trends in the burial data to help us understand these ancient people. Our discoveries help unite families as we learn more about children of God from all parts of the world.

Our analysis took two forms: a "supervised" portion and an "unsupervised" portion. The former sought to build a model capable of determining the sex of a mummy based on burial characteristics while the latter analyzed existing data looking for statistically significant correlations. We will expound upon these finding in depth below:

No. 01 — Supervised Findings

Beginning our review of our supervised findings, our focus must first and foremost be addressed. Specifically, we directed our efforts toward analyzing head direction. After various attempts with different models, we discovered that the Random Forest model produced the highest accuracy overall, achieving an impressive 81.68%.

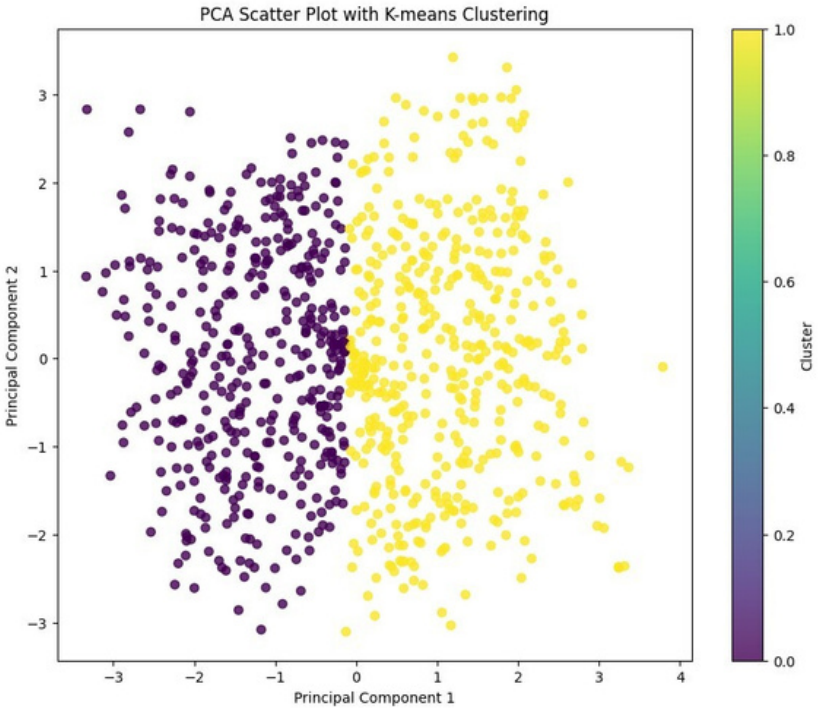
While we also experimented with a Decision Tree model, which yielded a respectable accuracy of 77.72%, and a Logistic Regression model, which attained an accuracy of 79.21%, it was the Random Forest model that truly stood out. Though this model does require a longer running time compared to the others, the increase in accuracy that it provides is undoubtedly worth the wait.

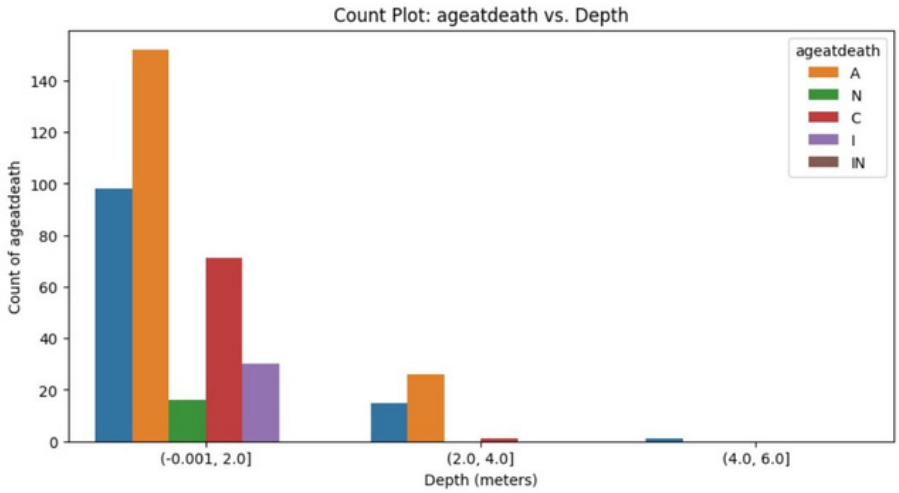
No. 02 — Unsupervised Findings

Let's delve into the exciting realm of our unsupervised model findings. Our approach to this analysis was markedly more exploratory than our previous methods. With an open mind and an eagerness to uncover new insights, we initiated our data sorting process and began by conducting a PCA analysis. This method proved to be particularly enlightening, as it enabled us to compare how features changed with depth. Buoyed by this initial success, we proceeded to perform a K-means analysis, utilizing the PCA background we had previously employed.

This approach allowed us to visualize two distinct clusters on the PCA Scatter Plot (shown below), an intriguing outcome that, while not as specific as we had hoped, served as a valuable starting point. Further examination of the PCA features themselves through looking at loadings revealed some fascinating correlations, but we were still in search of a more effective means of visualizing our findings. Enter the direct feature vs. depth comparisons via count plot, which ultimately yielded our most compelling results. Of these, perhaps the most intriguing was the discovery that no newborns or infants had been buried more than two meters deep in the ground (also shown below). This striking finding indicates that the area did not bury any infants or newborns until a later time.

Another intriguing observation was the correlation between hair color and depth. It appears that the closer to the surface you are, the greater the variety of hair colors you will find. Lastly, our analysis revealed that a mummy was more likely to be nothing more than bones, with no remaining wrappings, the deeper you dug. Overall, our unsupervised analysis was thought-provoking, providing us with valuable insights that shed light on trends within the area.



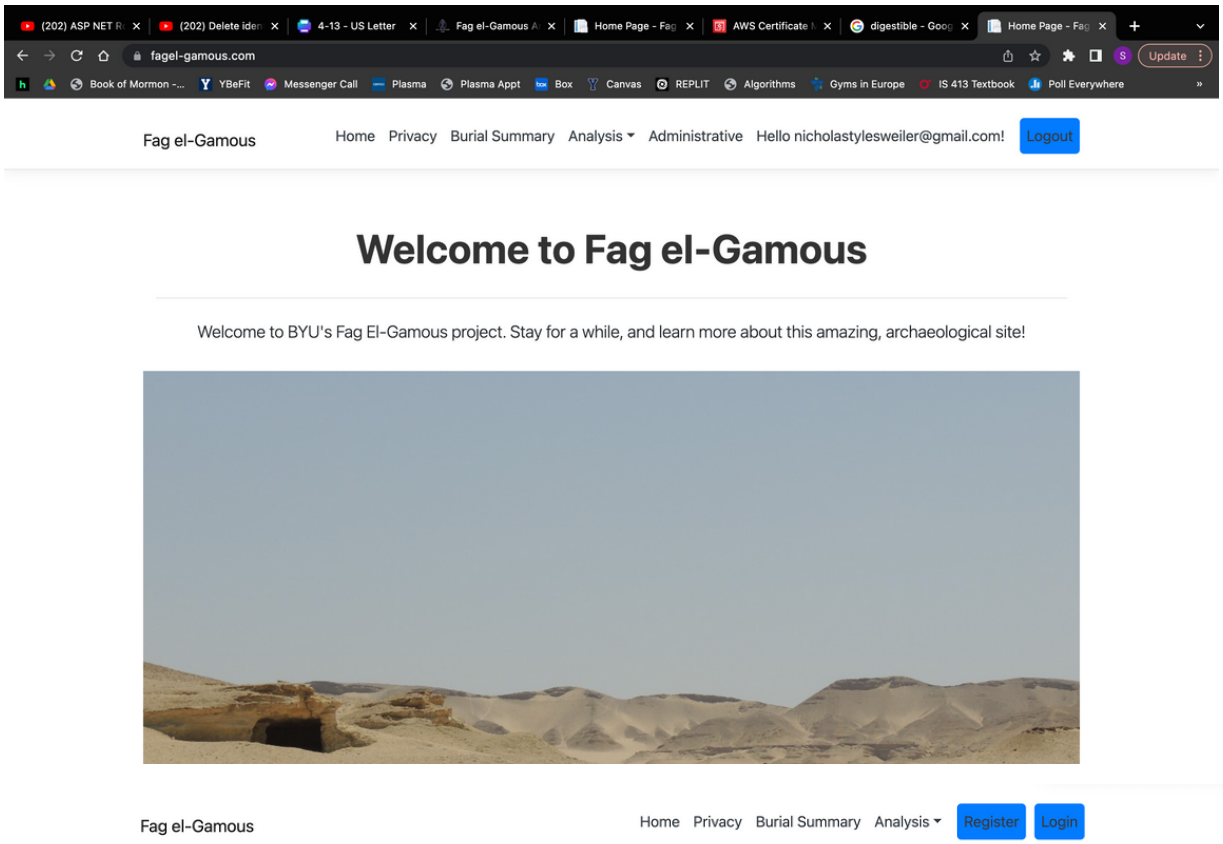


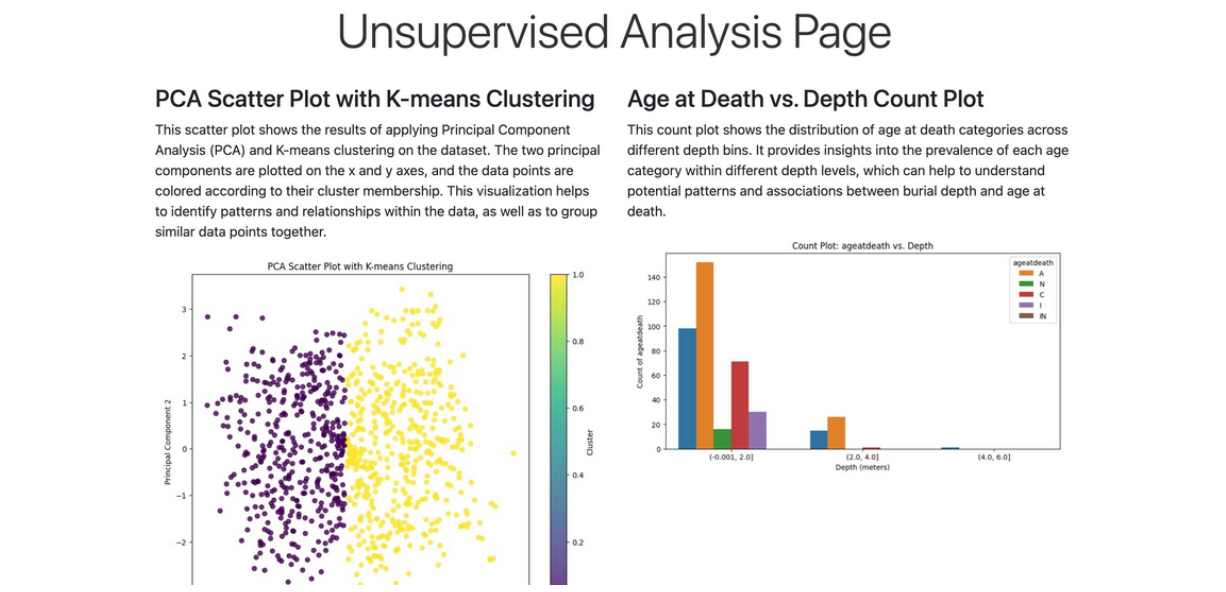
Our findings provide valuable insights into the burial practices of ancient Egyptians. We discovered that burial practices varied widely across different depths or time periods and that burial depths were often associated with particular attributes or groups. We also found evidence of changing burial practices over time, as well as the influence of changing genetics over time.

Overall, our project demonstrates the power of data analytics in uncovering insights and patterns in complex datasets. By combining supervised and unsupervised learning techniques, we were able to develop a deep understanding of burial practices in ancient Egypt and contribute to the ongoing study of this fascinating subject.

WEBSITE TOUR

As for our application, we have create a full-stack application designed in .NET Core 3.1 to integrate security and style. We will provide a basic tour of the site through images below. We encourage the reader to visit the site by clicking here. A more in-depth walkthrough will occur in the 413 section below.



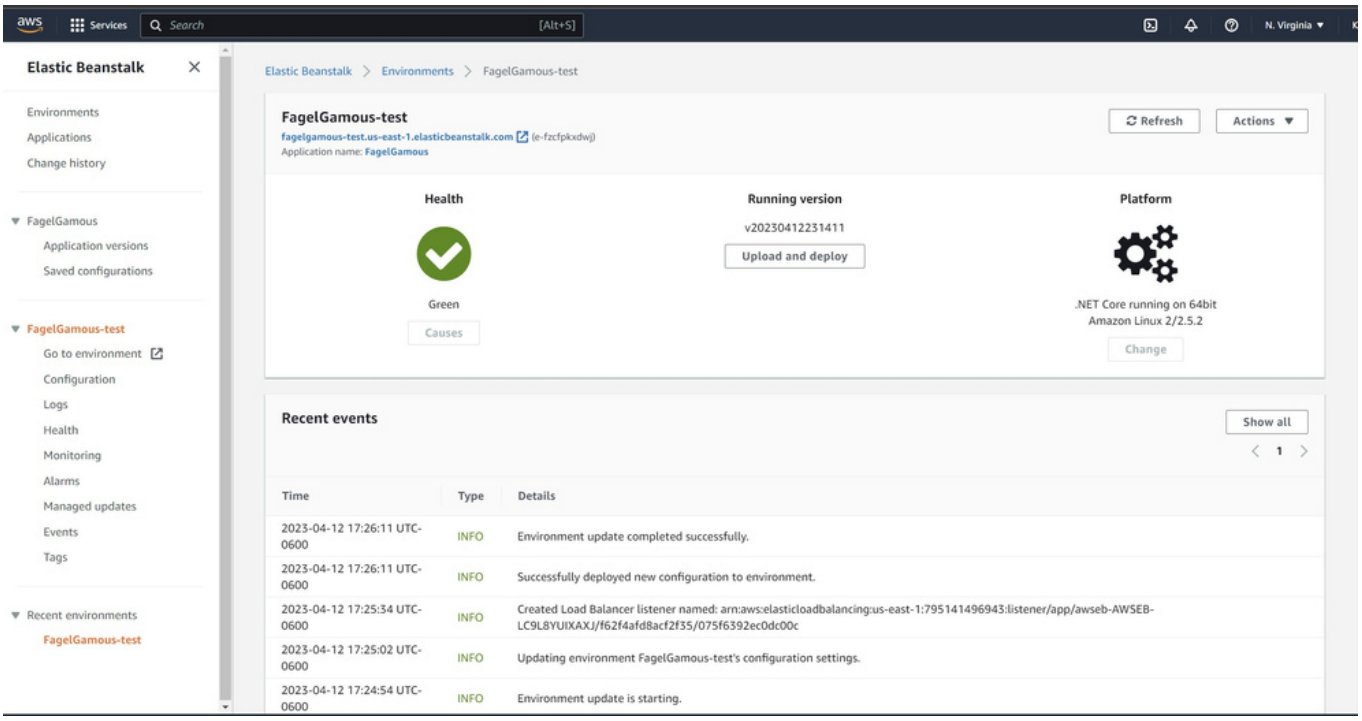


DOCUMENTATION

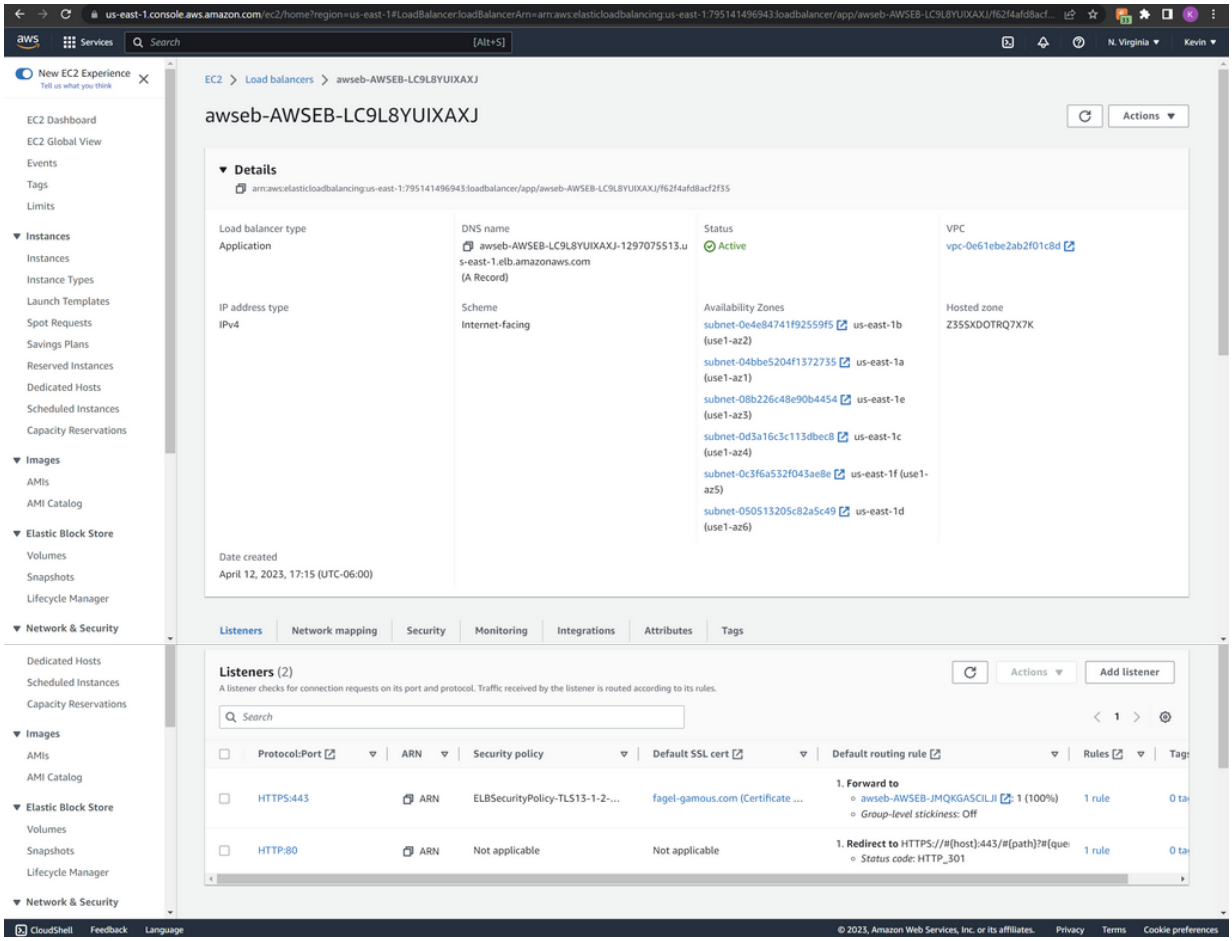
IS404 SPECIFICATIONS

No. 01 — Description of your existing solution

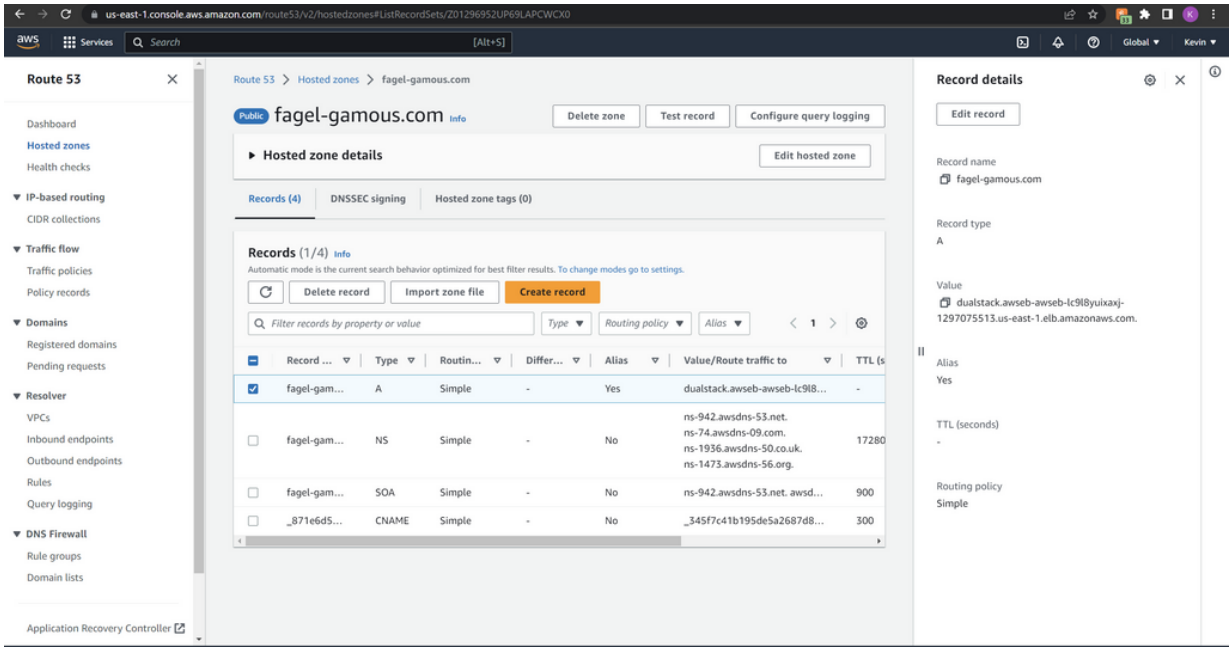
Elastic Beanstalk is a fully-managed service offered by Amazon Web Services (AWS) that simplifies the process of deploying, scaling, and managing web applications. With Elastic Beanstalk, developers can upload their code and let AWS handle the deployment, provisioning, and maintenance of the underlying infrastructure.



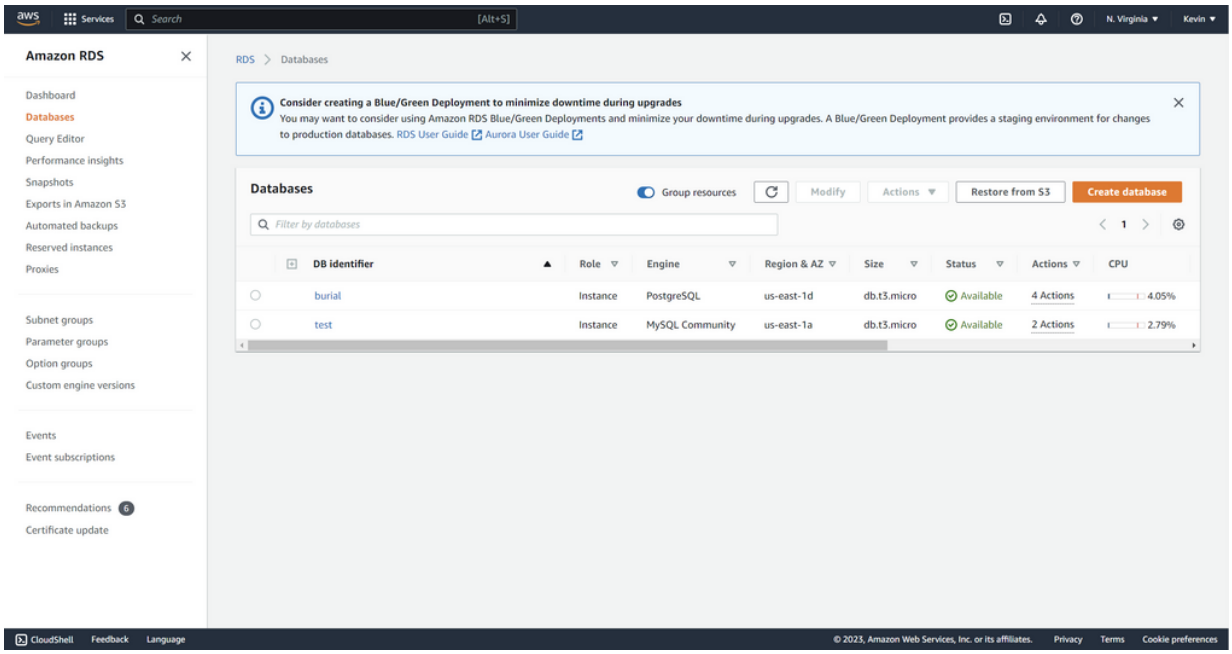
An Application Load Balancer (ALB) is a service offered by Amazon Web Services (AWS) that routes traffic to multiple targets, such as EC2 instances or containers, based on application-level routing rules. When used in conjunction with Elastic Beanstalk, ALB provides an additional layer of scalability, availability, and security for web applications.



Amazon Route 53 is a highly-available and scalable domain name system (DNS) service offered by Amazon Web Services (AWS).



Amazon Relational Database Service (RDS) is a fully-managed service offered by Amazon Web Services (AWS) that makes it easy to set up, operate, and scale a relational database in the cloud.

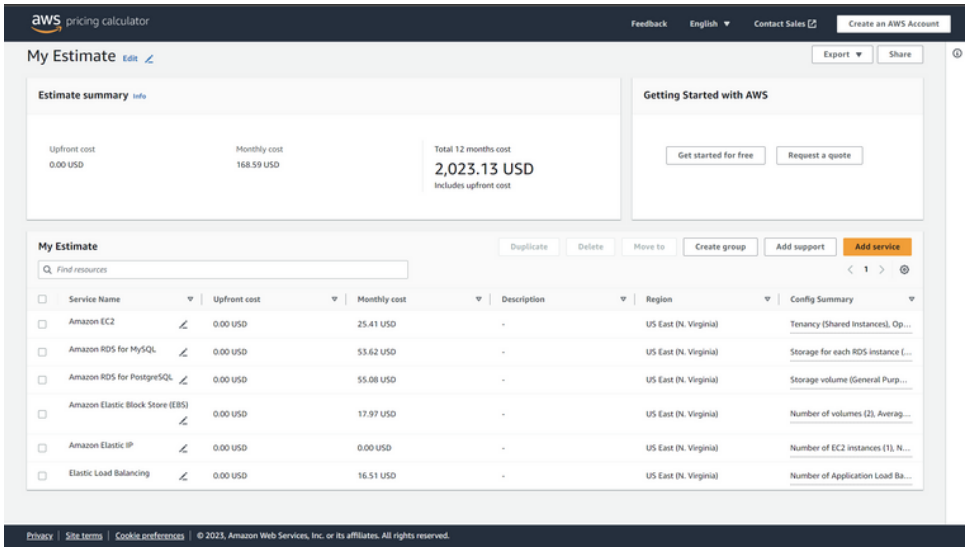


No. 02 — Service recommendations

1. AWS Security Manager: Security is paramount when dealing with sensitive information, such as user credentials and application data. AWS Security Manager provides a comprehensive and centralized approach to managing security across your AWS accounts and applications. This service enables you to ensure secure handling of user credentials, AWS endpoints, and other sensitive information while maintaining compliance with industry best practices and regulatory standards. By utilizing AWS Security Manager, you can minimize the risk of unauthorized access and data breaches, thus safeguarding your application and user data.
2. CloudFront: Delivering a fast and seamless user experience is crucial for any application. AWS CloudFront, with its edge locations throughout the world, including Egypt, can help you achieve this. By caching and delivering your application content closer to your users, CloudFront reduces latency and increases the speed of your application. This results in an improved user experience and a higher level of user engagement, which is particularly beneficial for a data-intensive application like yours that relies on real-time analysis of burial data.
3. Amazon S3: Efficiently storing and accessing the static content of your application, such as images, JavaScript, and CSS, is essential for optimal performance. Amazon S3 offers highly durable, scalable, and available storage that can be easily integrated into your .NET program. By storing your static content in an S3 bucket, you can ensure quick accessibility, better application performance, and reduced server load. Furthermore, S3's pay-as-you-go pricing model allows you to cost-effectively manage your storage needs as your application grows.

No. 03 — Estimated budget

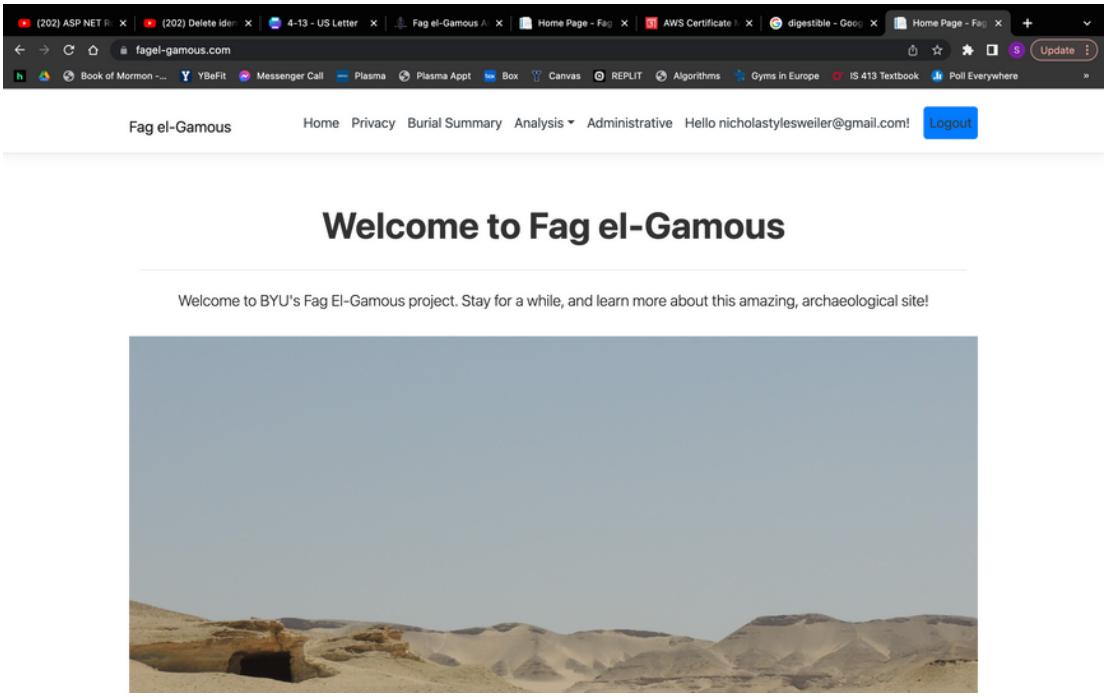
- The monthly cost estimate of hosting on a development server rather than free tier is $168.59.
- This estimate includes using current technologies as well as those prescribed above in the recommendations section



IS413 SPECIFICATIONS

No. 01 — **Home Page**

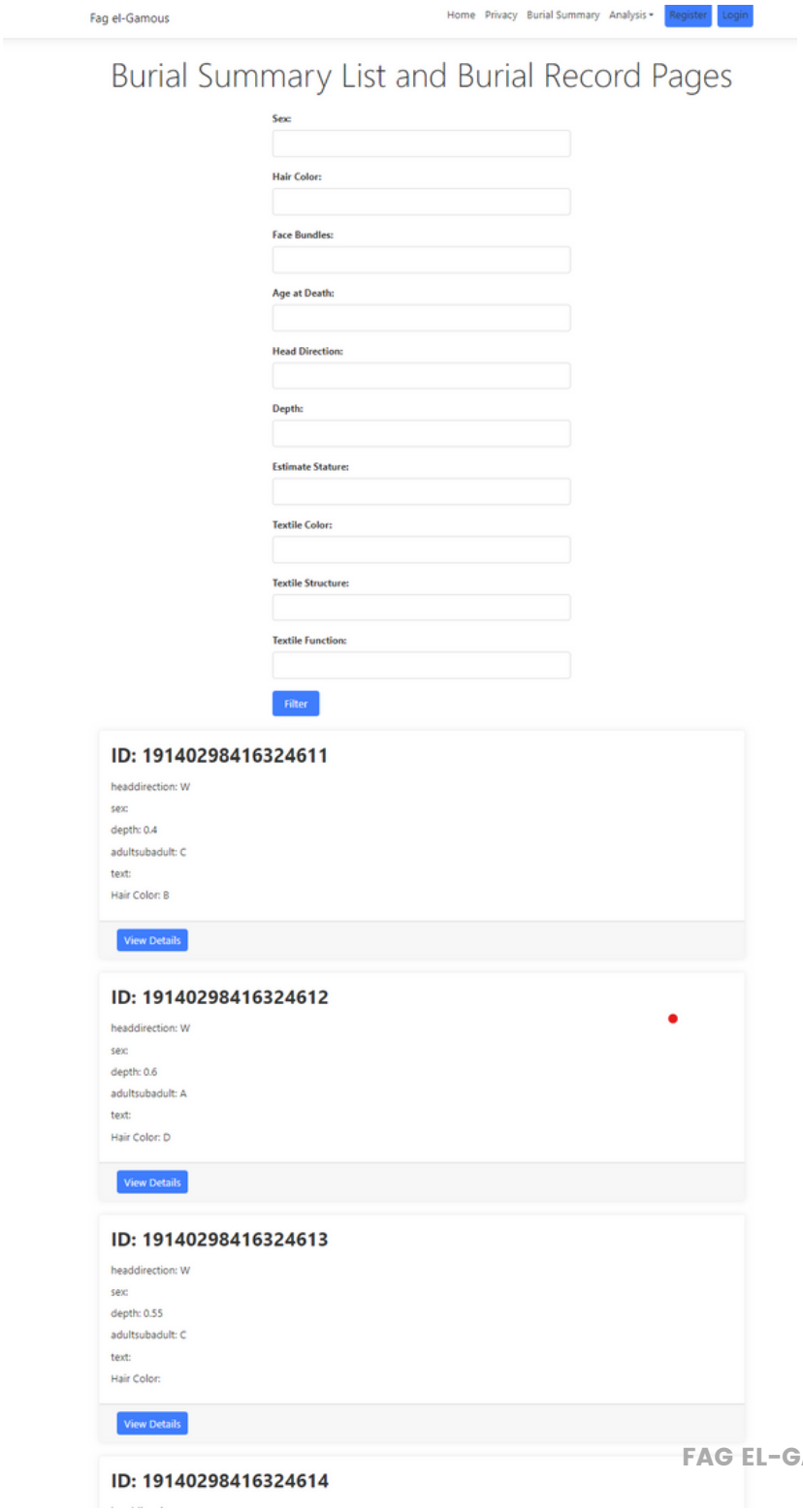
- The link to view the burial information is accessible via the Burial Summary navigation tab.



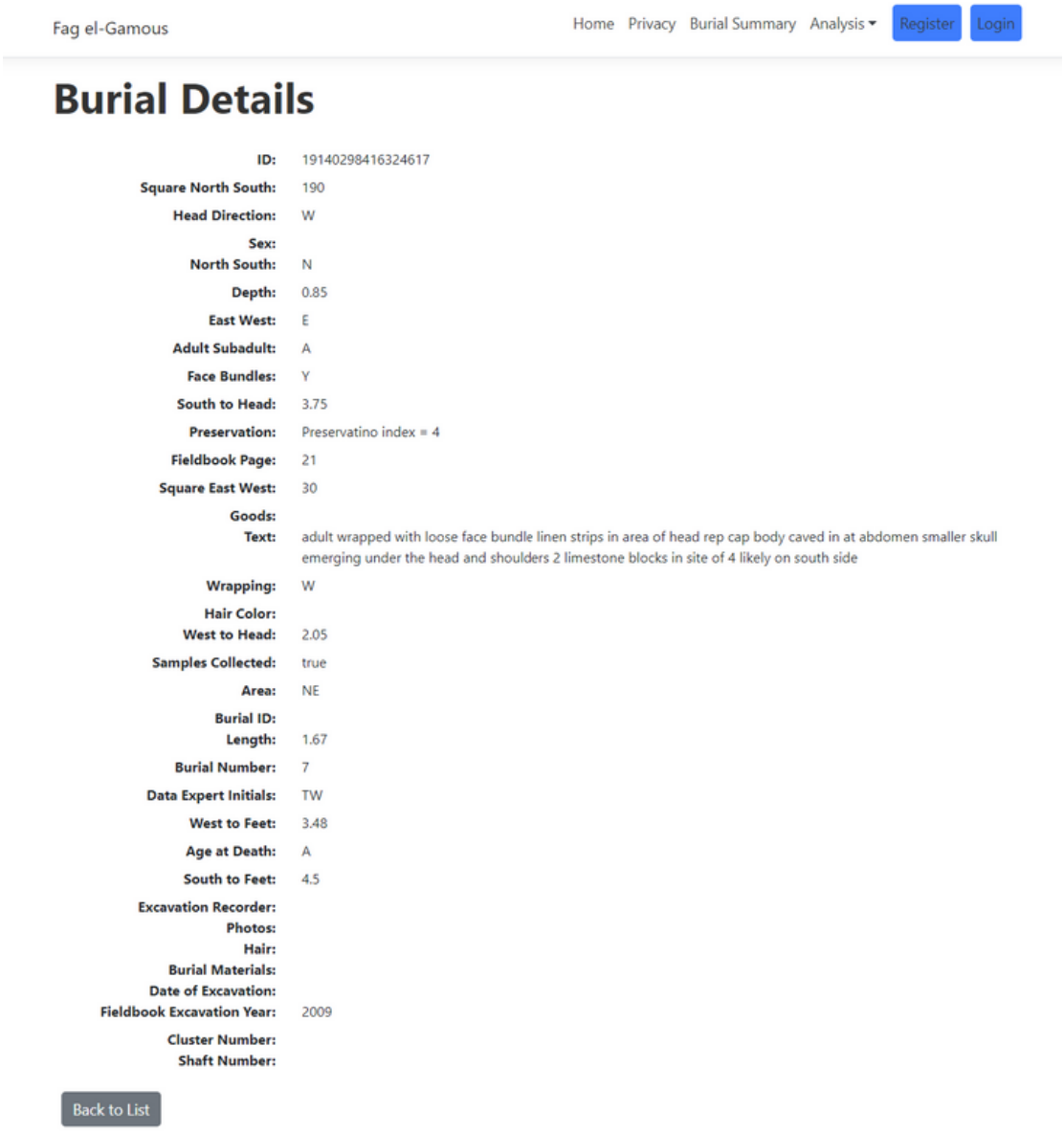
ProjectsViewModel.cs | Burial.cshtml | CreateBurialViewModel.cs | HomeController.cs | Index.cshtml

```
@{
    ViewData["Title"] = "Home Page";
}


<main role="main" class="pb-3">
    <div class="container">
        <div class="row">
            <div class="col text-center">
                <h1 class="mt-5">Welcome to Fag el-Gamous</h1>
                <hr class="my-4">
                <p class="lead">Welcome to BYU's Fag El-Gamous project. Stay for a while, and learn more about this amaz
            </div>
        </div>
    </div>
</main>

<div class="container-fluid px-0">
    <img class="img-fluid" src="~/img/egypt.jpeg" alt="Egypt" />
</div>
```

- A paginated list of all burials with a brief summary of the info under the Burial Summary Page
- The ability to filter the results by certain criteria based on client requests (10 most commonly)

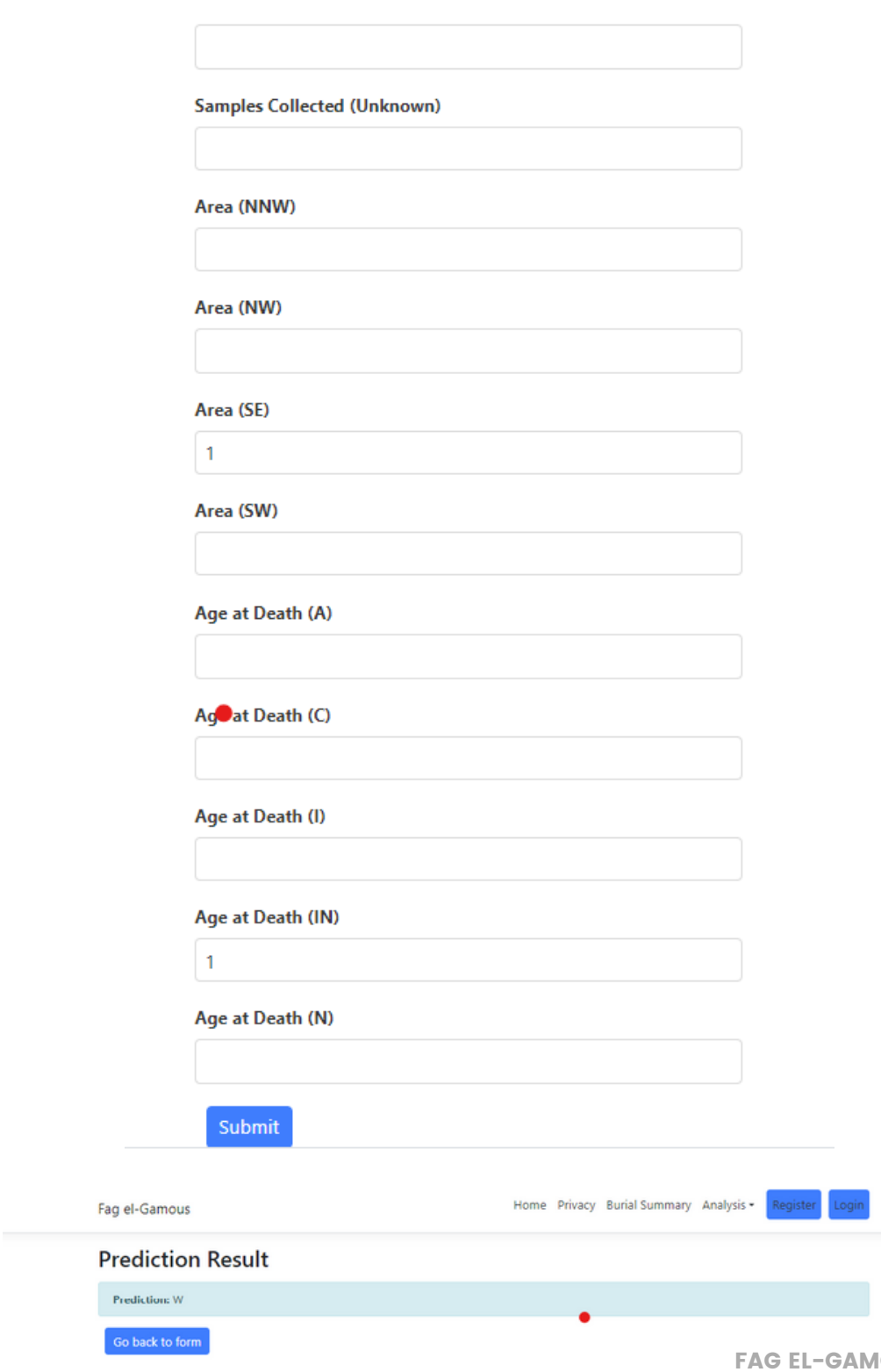


- The ability to view an individual record by selecting the "View Details" button.



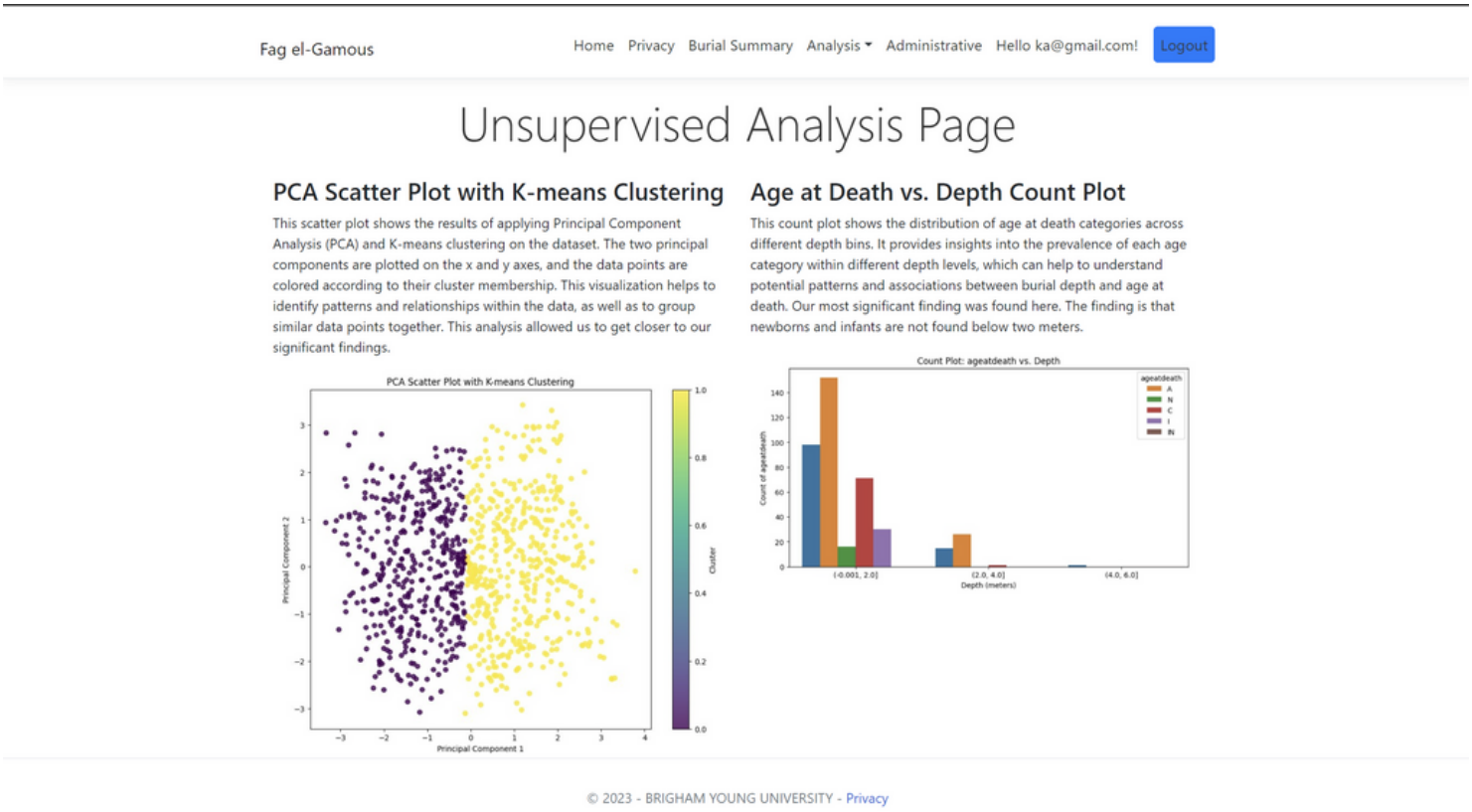
No. 03 — Supervised Analysis Page

- A page that allows the user to enter osteology and/or burial information to predict (based on your trained ML model) the sex of the burial (see information re: IS 455 requirements)



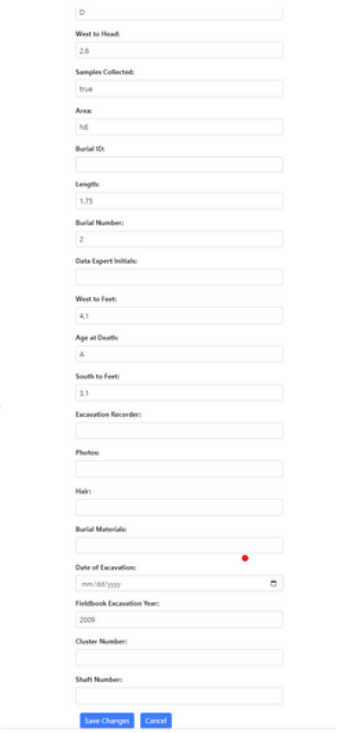
No. 04 — **Unsupervised Analysis Page**

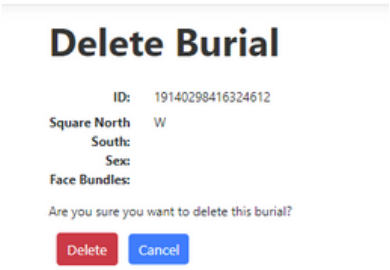
- A page that displays the results of your unsupervised analysis(es) for the client (see information re: IS 455 requirements)
- Included with visualizations by selecting Analysis and then Unsupervised in the dropdown menu.
- Note that these do not need to be live visualizations, but can simply be static images that you are displaying on the page with the accompanying explanation(s)

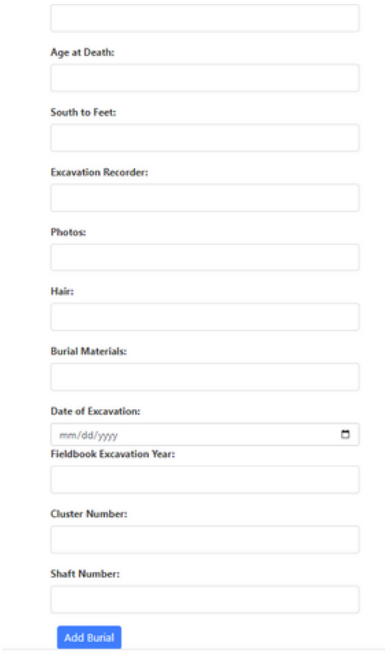


No. 05 — **Administrative Pages**

- Ability to add/edit/delete records (authorized users have CRUD buttons associated with the Burial Summary page views for each record
- Ability to manage user accounts (using native Core Identity in the Administration section)

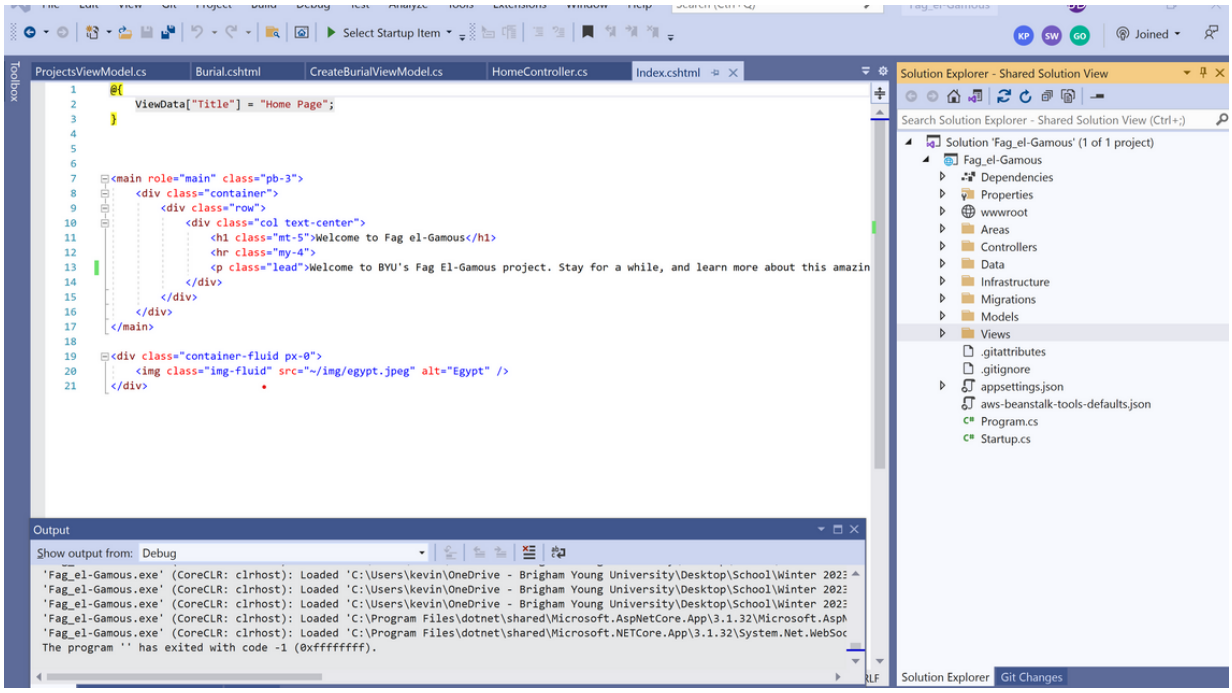






No. 05 — **Technology Stack**

- ASP.NET Core MVC (using version 3.1)
- AWS (Elastic Beanstalk for deployment)
- EC2 for Web Server (configured through Elastic Beanstalk)
- RDS for database (See AWS section of 404 for more details on all of these)

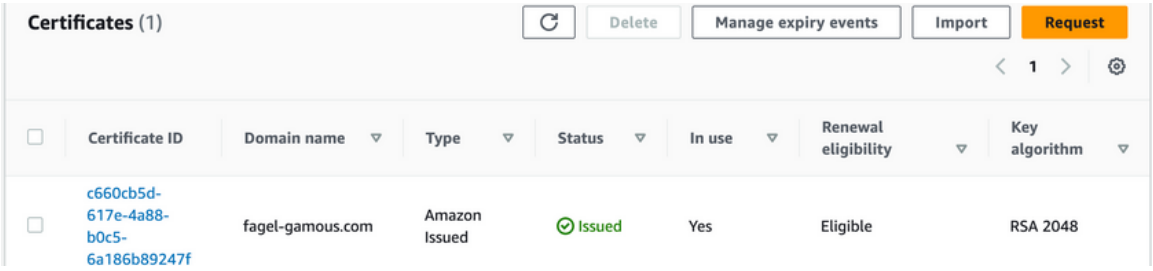


IS414 SPECIFICATIONS

Implementing such functionality would be in the best interest of Larson Davis because it would streamline bug resolution, saving the firm money and labor hours. The project will improve the company in the following areas:

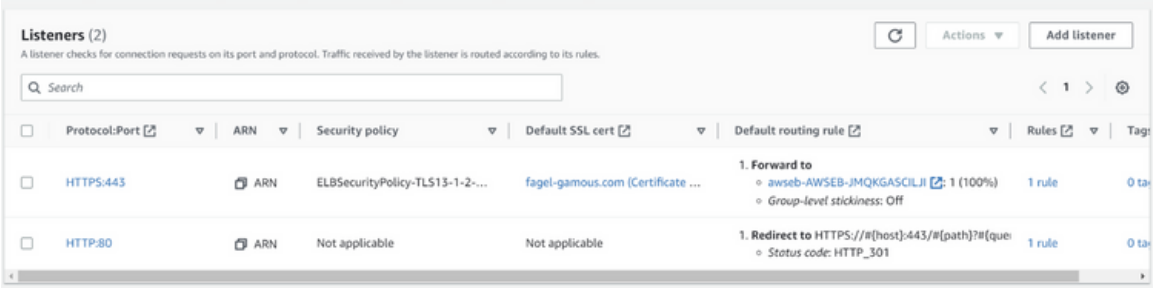
No. 01 — Encryption

- HTTPS was encrypted for all public traffic using a TLS certificate through AWS Certificate Authority.



- HSTS was enabled for our site using the internal .NET command "useHSTS()" in our Startup.cs file.

- All HTTP traffic was routed to HTTPS through Application Load Balancer Listeners.



No. 02 — Authentication

- Authentication for users (username/password) was using ASP.NET Identity, along with its native login UI.

- Configuration of ASP.NET Identity to require passwords with at least 12 characters as well as numbers, special characters, and uppercase characters were specified in the ConfigureServices class of the Startup file.

- Two-Factor Authentication (2FA) was satisfied through implementation of Google Authenticator. Users are provided with a TOTP through a smartphone which is requested upon successfully inputting their username and password.

No. 03 — Authorization

- Role-Based Access Control (RBAC) exists as two types of users: authorized and non-authorized. Due to the needs of the client, no other roles were required of the application. The non-authorized user is required to login before performing CRUD operations on the Admin page by using the Identity Core's built-in "IsAuthenticated" method.

No. 04 — Integrity

- In a similar vein as Authorization, non-authorized users cannot access the Admin page to perform CRUD on users, nor do the "Edit" or "Delete" buttons render for burial records because of the "IsAuthenticated" method.

No. 05 — Credentials

- Due to time constraints, this requirement was not fully fleshed out. However, our team plans to use AWS Secrets Manager to handle the ConnectionStrings and other credentials in the application.

No. 06 — Other

- A GDPR-complaint privacy policy was created specifically for the Fag el-Gamous project and a link can be found at the footer of each page which navigates to the Privacy page of the site.
- A GDPR-complaint cookie pop-up occurs on load for every user.
- Compliance with the Egypt Data Protection Law passed in 2020 requires recommendations for clients to properly obey the law. We will include these below:

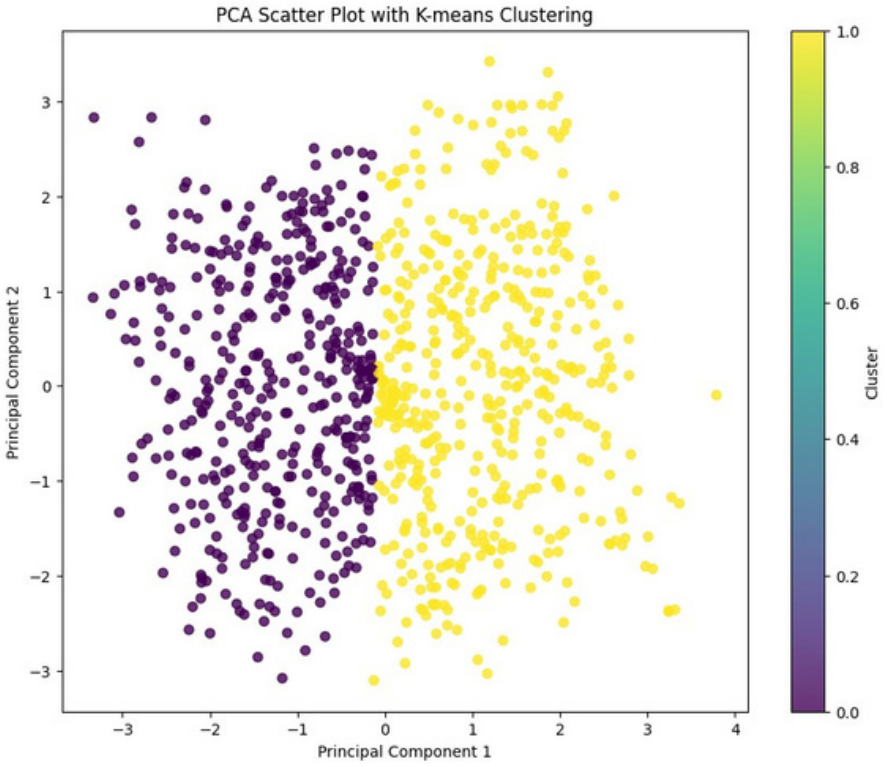
- Appoint a Data Protection Officer and establish a robust data protection policy
- Obtain consent from users before collecting their personal data. Provide transparent privacy notices on your website
- Implement mechanisms for users to exercise their rights under the law, including the right to access, rectify, erase, and object to the processing of their personal data.
- Report data breaches to the Egyptian Personal Data Protection Center within 72 hours of discovery.
- If transferring personal data outside of Egypt, implement appropriate safeguards such as contractual clauses or binding corporate rules.
- Maintain detailed records of data processing activities.

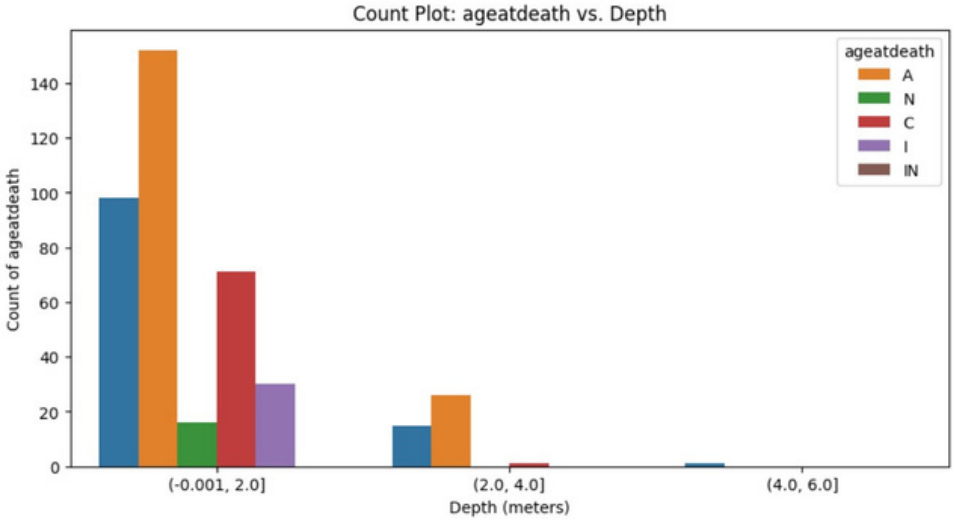
IS455 SPECIFICATIONS

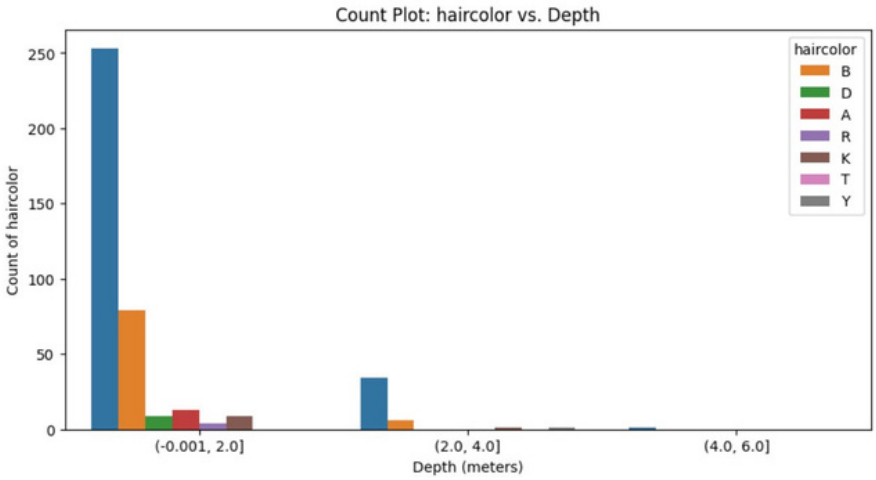
No. 01 — Building Supervised and Unsupervised Models

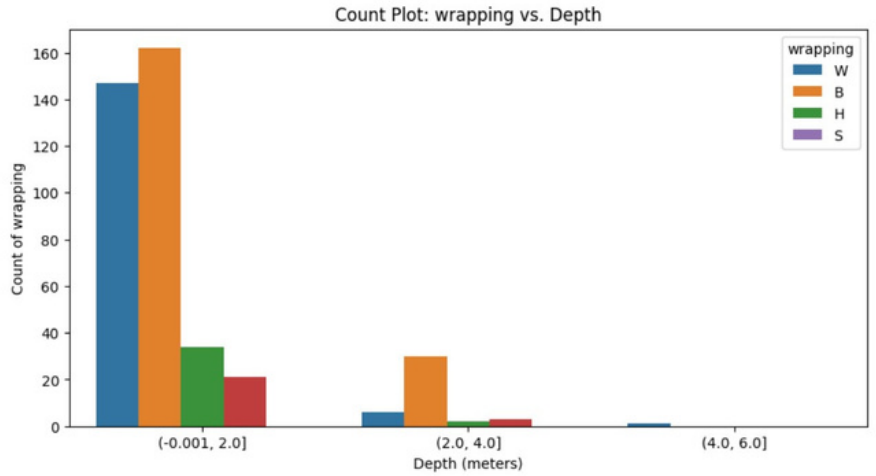
We created two .ipynb files that are linked above. One is for a supervised model predicting the head direction and the other is an unsupervised model seeing how features change at different depths. We used the provided data (the fagelgamous_database.backup) to create the models. For both models, we imported the data we felt was useful, chose which columns we wanted to use, and then further cleaned it. Now let's get into model specifics. For the supervised model we ran three different models to try and find the best prediction model we could. We used the Decision Tree, Logistic Regression, and Random Forest Models. The Random Forest Model far exceeded expectations and gave us an 81.68% accuracy percentage. This made us decide to use the Random Forest Model for our onnx file export.

Now on the unsupervised model, we had layers of analysis that built up to finding the significant data we arrived at for our conclusion. We ran a PCA analysis and then furthered t that by applying k-means clustering. We referenced the loadings for each PCA analysis and were able to find a correlation in the data but we wanted better visuals. The visuals we settled on were scatter plots of numerical features vs depth and count plots for categorical features vs depth. We found the relationships of age at death, hair color, and wrapping with depth to be significant because of noticeable trends. Below is the PCA Scatter plot with K-means analysis and the significant count plots that show the body of our work on the unsupervised portion.









We specifically found that the lack of infants and newborns below two meters in the age-at-death count plot, the new hair colors that emerged closer to the surface in the hair color count plot, and the significant difference in the frequency of wrapping being present closer to the surface as opposed to deeper in the wrapping count plot to be significant.

No. 02 — Deploy the Supervised Model

For deploying our supervised model, we used an ONNX file that had our Random Forest Model on it and deployed it through an ASP.NET API. From there we tested it with Postman and found it worked, and then created the ability for a user of our website to submit a form using those same features and get back the head-direction prediction. Here is the GitHub repository for the ASP.NET API so you can see that code as well: https://github.com/gavin3567/BurialAPI.git