INTEX WINTER 2023 - Fag el-Gamous

Fag el-Gamous (which means Way of the Water Buffalo) is a large and important cemetery in Egypt that has been under archeological investigation (primarily by BYU researchers) for over 25 years. The excavation has garnered international attention because of the interesting insights gained by studying the Fag el-Gamous mummies. Detailed background information on this site can be found in the academic article "Rethinking burial dates at a Graeco-Roman Cemetery: Fag el-Gamous, Fayoum, Egypt."

During multiple years of study, researchers have excavated more than 1,000 mummies and collected data in field notes, spreadsheets, and simple databases about each mummy's location, physical orientation in the ground, burial depth, hair color, wrapping (including information about the textiles used), sex, age, bone (including skull and teeth) dimensions, items accompanying the mummy, and more. And, of course, there are many, many photographs. Offsite, researchers have performed a variety of analyses on the excavated items including carbon dating, DNA sequencing, and others.

The "dig" is organized using a grid system where a square is defined by the coordinates North (N), East (E), South (S), and West (W) of an origin. It is common for two numbers to designate the distance from the origin in a direction (e.g., N 190/200 or E 40/50). Those squares are then subdivided into the Northwest (NW), Northeast (NE), Southeast (SE) or Southwest (SW) quadrant (section) of the square. An example burial location is: **N 190/200 E 40/50 SE**. Because many mummies (usually 40 or more) are found in each location, a number is assigned (e.g., **N 190/200 E 40/50 SE #9**) to uniquely identify the burial.

As the quantity and diversity of data associated with this excavation have grown, a few challenges (and opportunities) have emerged. The researchers would like to engage your team to complete a project that has 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.

IS students previously cleaned and organized the data so that it is currently stored in a PostgreSQL database.

Data

The data for the project is provided at the following link:
- **https://byu.box.com/s/3zj5ezth7firwo3qn9dvv51jtyq1s6ht**

This data has been collected across decades of investigation by multiple researchers and assistants. You have the great fortune that this data is largely cleaned and already in a format that can be quickly inserted into a

database. The clients are currently using PostgreSQL and the data is in that format. There are also images of burials, artifacts, and textiles that could be used in your site or in your analytics should you choose. Included with the data is a current ERD of the database to help you understand the structure of the data. Other information about the tables that might be helpful:

BurialMain - This is the primary table in the database and represents the excavation of a single person.

Textile Tables
Many tables in the database store information about textiles (clothing, ribbons, etc.) found with the burials. These include:
Textile - The main textile table
TextileFunction - The function of the textile as determined by the researchers
Color - Colors of the textile
Structure - Textile structure
Dimension - Textile measurements
Decoration - Any decorations on the textiles
Analysis - Results of any tests done on the textiles after excavation
YarnManipulation - Information about the type of yarn/thread/weave that was used to create the textile
PhotoData - Information about textile images

Bone Tables
Two tables in the database store measurements made of bones.These include:
BodyAnalysisChart - Measurements of important bones for determining sex, height, etc. that are not the skull.
Cranium - Skull measurements

Other Tables
Artifacts - information about pottery, toys, jewelry, and other artifacts found with a burial
Biological - Test results that have been performed after excavation
C14 - Results of carbon-dating tests (limited or missing)

Users

There are two primary types of users who will interact with this site:

1) *Members of the Public -* These users will not need to have an account or authenticate to gain access to the information. One goal of this project is to share this data with the world. These users will be able to view, but not edit data.

2) *Researchers -* BYU researchers (and colleagues) will need user accounts with authentication to enter data into the system (whether located at Fag el-Gamous or elsewhere), and edit/delete the data.

If you choose, you might create additional roles (e.g., an administrator role who is a superuser authorized to perform all functions within the system including those of researchers, but also other maintenance operations). But this is not required.

Course-Specific Instructions

Enterprise Application Development (IS 413)

Home Page
- A professional-looking landing page that describes the project and has a link to view the burial information

Burial Summary List and Burial Record Pages
- A paginated list of all burials with a brief summary of the info.
- The ability to filter the results by certain criteria
  - The database has many, many fields, and we do not expect you to be able to filter by all
  - The client will have some insight into which filter options should take priority
- The ability to select a specific record and pull up the details of the burial (i.e. field notes, photographs, measurements, and any details captured about the mummy or items included in its burial.)

Supervised Analysis Page
- A page that allows the user to enter osteology and/or burial information to predict (based on your trained model) the sex of the burial (see information re: IS 455 requirements)
- One option you may consider in your analysis is to use something like this example to build a ML Model that utilizes an .onnx file in ASP.NET:
  https://towardsdatascience.com/deploy-sci-kit-learn-models-in-net-core-applications-90e24e572f64
  - (Note that this example uses the ASP.NET API template, but could be easily adapted to instead pull the input information from a form to pass to the HTTP Post method on the controller.)

Unsupervised Analysis Page
- A page that displays the results of your unsupervised analysis(es) for the client (see information re: IS 455 requirements)
- Include your visualization(s)
  - 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)

Administrative Pages
- Ability to add/edit/delete records
- Ability to manage user accounts

Technology Stack
- ASP.NET Core MVC
- AWS
  - EC2 for Web Server (may be configured through Elastic Beanstalk)
  - RDS for Database is recommended, but other options may be considered. Remember the limitations of SQLite discussed by Professor Schuetzler in 404.

Bonus Functionality / Stretch Goal
- React Front End

Infrastructure (IS 404)

The important IS 404-related concepts from INTEX come into play in the deployment of your application. Remember the things we have discussed related to why people use cloud computing, and why AWS is able to charge money for the services they provide (like EC2, RDS, etc.). Companies use AWS for the value they provide, and you should be using the appropriate AWS services and settings for the product you are creating.

Your application should be deployed on AWS using the appropriate services. In addition, your written document should include a list of services used, and recommendations for any services not implemented in the current iteration. This should also include a monthly cost estimate for running the application.

> **We highly recommend you use a personal AWS account to do your deployment.** The limitations of the AWS Academy lab environment will cause you unnecessary strife during your INTEX deployments. The AWS Free Tier (for new accounts) includes some level of free EC2, RDS, S3, data transfer, and many other things. If you are judicious in your usage, you should be able to complete Intex with minimal-to-no costs. Once INTEX is over and grading is completed, you should stop or delete your services to avoid continued charges.

Requirements

- Deploy your application to AWS, including a database.
- Have a custom domain. You can either purchase your own (recommended) or use a subdomain of is404.net. The Beanstalk-provided URL does not count as custom.
- You do not have to use AWS services exclusively, but the primary computing and database need to be there.
- Your website should be accessible through HTTPS.
- Select the appropriate AWS resources for each service needed to complete the requirements of the project. Document any choices and services used in the deliverable.

Include in the Deliverable

- ☐ Description and documentation of your existing solution - what (AWS or other) technologies have you included to build and deploy your current application.
- ☐ Service recommendations - If there are services that you have not yet implemented that would be beneficial, include recommendations with an explanation for why it should be used.
- ☐ Estimated budget - estimate the monthly budget to provide this application, assuming ~100 daily users.

Machine Learning (IS 455)

We have been building our 'ML tool boxes' throughout the semester with a variety of supervised and unsupervised ML techniques. The data for this project provides many opportunities to employ several of these techniques to help the Fag el-Gamous research team answer questions (some they know they have, and some they have never thought of).

One important issue the research team deals with on a very regular basis is to determine whether a buried person was female or male. Currently, information using bone measurements (i.e., osteology) is used to make this determination. What has not been used in the sex determination are factors related to burial itself. For example, were ribbons found with this person? In which orientation was the head placed? What types of textile (e.g., weaving patterns) were found with this person? Numerous possibilities such as these have never been explored alongside the osteology data.

In addition to the issue above, the research team has a working hypothesis that the burial practices in the 1,000-year-old cemetery could have changed over time. The earliest burials have been found at a depth around 3-4 meters, with the more recent burials found much closer to the surface. While many insights could be uncovered with the data, one possibility is that textiles of one color were more prominent at earlier times. Another possibility is that burial orientations changed over time or certain types of people were buried at deeper depths than others. We simply do not know what will be found in exploring this working hypothesis.

Requirements

- **Building Supervised and Unsupervised Models**
  - Create an .ipynb file that uses the provided data to create at least one supervised and at least one unsupervised ML model.
  - For the **supervised** portion, you will help the research team with the sex determination issue mentioned above. Using the osteology and burial data together (something that has never been done), create a model to predict whether the buried person was sex female or male.
  - For the **unsupervised** portion, you will try to provide insight into the working hypothesis mentioned above. Specifically, can you and your team uncover any important details in the data that would help the research team support/refute the argument that the burial practices changed over time?
  - Meaningful visualization(s) should accompany the **unsupervised** model(s).
  - Document your data cleaning, data imputation, feature selection, validation, and algorithm selection process so that TAs can review your .ipynb file and easily understand your work.
  - You may train other models as well if you have an idea you want to implement. However, this is not a requirement.
- **Deploy the Supervised Model**
  - You may (1) save your model as a pickle or ONNX file and store it in your web application or (2) create an API that uses your trained model that you can call from your web application.
  - As stated in the web application description, this model will be used to inform the research team whether the buried person that they have gathered and entered information about is sex female or male.
- **Not required**
  - You do not need to set up your environment to automatically retrain your model.

- You do not need to allow the ability to retrain your model(s) should new records be added to the database.
- NOTE: both of these features above are essential parts of a machine learning environment in industry. They are not required to keep this project from getting overly complex.

Security, Authentication, and Authorization (IS 414)

The IS414 portion of INTEX includes BOTH applying skills you have already learned AND learning how to implement a few new skills now that you have sufficient technical background to complete them.

Requirements

- ☐ Encryption
    - ☐ Use HTTPS for all public connections. It is fine if you use a subdomain, but you must have a valid certificate to enable TLS.
    - ☐ Enable HTTP Secure Transport Security (HSTS) on your site. This can be tricky or trivial depending on your chosen cloud architecture. A single line of code is sufficient in some cases, but NOT in others. Don't underestimate the time required here.
    - ☐ Redirect HTTP traffic to HTTPS.
- ☐ Authentication
    - ☐ Provide the functionality to authenticate users using a username/password (likely using ASP.NET Identity).
    - ☐ Configure ASP.NET Identity (or your chosen Identity provider) to require better passwords than the default PasswordOptions use (see https://docs.microsoft.com/en-us/aspnet/core/security/authentication/identity-configuration?view =aspnetcore-3.1). You should know what "better" means.
    - ☐ Enable at least one form of two-factor or multi-factor authentication. You have flexibility in the choice and implementation. Note: You must have some accounts without 2FA or MFA for grading purposes so that the TAs can access your Website for grading without having your phones or faces.
- ☐ Authorization
    - ☐ Use Role-Based Access Control (RBAC) in a manner that only an authenticated user has the ability to add, modify, or in rare cases delete data. You may choose to have additional roles beyond an authenticated user and non-authenticated user, but this is not required.
- ☐ Integrity
    - ☐ Data should only be able to be changed or deleted by an authenticated user and there must be confirmation required to delete data.
- ☐ Credentials
    - ☐ Handle credentials (usernames and passwords, API keys, etc.) safely. You may choose to use a secrets manager, a separate file that is not uploaded to a code repository like Github, or an environmental variable. But you should not include credentials in your code. Please note that having a functional Website is worth more overall points than protecting credentials properly, so prioritize wisely (as much as this pains your instructor). Document this for the TAs if it isn't obvious how you are implementing this.
- ☐ Privacy
    - ☐ Create and populate the content of a GDPR-compliant privacy policy that is linked from the footer of your Web App (at a minimum on the home page). You may use existing templates (see https://gdpr.eu/privacy-notice/) or or generators (see https://www.privacypolicies.com/privacy-policy-generator/ or https://termly.io/resources/templates/gdpr-privacy-policy/), but the content should be tailored to your site.

- [ ] Enable a GDPR-compliant cookie consent notification
- [ ] Other
  - [ ] Enable the Content-Security-Policy (CSP) HTTP header. Specify the sources you need for your site to function and no more (e.g., you might choose to define default-src, style-src, img-src, script-src, etc., but only choose what you need).
  - [ ] In 2020, Egypt passed a Data Protection Law. Provide brief recommendations (bullets are fine, 2 paragraphs is the absolute maximum length) to the clients on what they need to know and do to comply with this law. Put this in the submission document. If you choose to implement something in your Web application to comply with this (not required), please document it.

Submission

Deliverable Submission

Deliverable Due Date: April 14 at 11:00 am (the morning of the presentations).
Deliverable Submission Location: The IS 414 (Security) Learning Suite course under the INTEX 2 assignment. Only one submission per group is needed.
Deliverable Name: Your document should be named your group number with the appropriate file extension. For example, if your group was 2-16 and you are submitting a PDF document you should submit the ***2-16.pdf*** file as your submission.

Written Project Description Design

You are to submit a document (e.g., PDF or Word) containing the following:

1. **Group Information**

	Group Information
Group Number	
Group Members	

2. **URLs and User Credentials -** Include completed versions of the following tables. Grading will go VERY badly if you make typos here.

	URL
Website URL	
WebsiteURL Backup (not required)	
.ipynb URL	
.ipynb URL (if more than one)	

	URL	GitHub Branch
Repository URL		

Note: <u>Please ensure when submitting that your GitHub repository is set to Public for grading purposes.</u> You may submit the URL to the ipynb in GitHub or Google Colab.

While you need to be very careful with credentials, please provide faculty and TAs the login credentials to your ASP.NET Core application (including users with different roles if implemented) so that they can properly grade your solution.

Username	Password	Role	2FA/MFA Used (if any)

Note: The TAs MUST be able to access the Website without needing a "something the team has" or "something the students on the team are". You may need to create some accounts that do not require MFA for this purpose.

> **Details of your solution (divided by class)** - While all (or almost all) of the grading will be completed by evaluating your Web application, code repository, and formal presentations, it can be useful to specify some of the details of your solution. This section of the document has no length requirements and may even be omitted if you are certain that your Web application **VISIBLY** meets all project requirements (e.g., deployment to AWS). The point of this section is to, without rambling, include information to convince those grading your projects that you successfully applied your skills to meet the requirements. If your solution to a requirement isn't going to be immediately obvious to graders, you should briefly mention it. You might also call out any ways that you went beyond the requirements. Again, there is no specific requirement for length or formatting of this section of the document other than it should be readable and divided by class.

Please follow the file naming conventions described above and include the necessary information in the tables above your document at the top of your document. The fastest way to upset your graders is making them do extra work to figure out which group submitted this document, where to find the correct URLs, or how to log in to your site. Maximize your score by keeping your graders happy.

3. **Executive Summary -** Each team should submit a 2-3 page executive summary that should include screenshots and a summary of your project, with particular emphasis on the findings from your supervised and unsupervised analyses.

Peer Eval

You will also need to complete the INTEX2 Peer Eval by **Friday April 14th at 11:59 PM**. The survey can be found here: https://byu.az1.qualtrics.com/jfe/form/SV_1Cide0O1PCsWFOm. Please complete this immediately after your presentation. The results of your eval will be kept confidential, but it will affect team member grades. We want this to be fair for those that did the work. The solution is to make sure that you pull your weight and more. Help the team and be a great team member helping where needed and doing quality work!

Presentation

Your submissions will be judged by panels composed of industry professionals, faculty, the client team, and possibly TAs. Your submissions will be evaluated based on the overall solution, presentation, development of your Website, security of your solution, and your infrastructure.

You will be randomly assigned a presentation slot. Presentations will be held Friday beginning at 12:00 pm. A schedule with the assigned times for each group will be posted. Each presentation period will be structured as follows:

Presentation and Tech Demo	20 minutes (Student group presents)
Questions & Answers	5 minutes (Student group answers questions)
Judge Deliberation	5 minutes (Student group waits in the hall)
Feedback	5-10 minutes (Student group comes back to room)
Break	5 minutes

The judges for your presentation will take the role of the clients. Your role is to pitch your system/solution as worthy of continued investment and/or adoption by the client. Get them excited about the system and how it meets their needs. The judging panel will be evaluating how your system works, how and whether it meets the client's needs, the quality of the presentation, and whether you creatively went beyond the minimum requirements. The judges have some technical background, but are primarily interested in the functionality, features, usability, feasibility/cost, and impact of your solution. The tech demo is likely the most important part of the presentation. Prioritize showing us your solution.

Do NOT bring handouts to the presentation.

Presentation Schedule
This schedule might have slight adjustments during INTEX week. Double check your room/time on presentation day.

	Judge Team										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
ROOM	TNRB 170	TNRB 180	TNRB 184	TNRB 210	TNRB 220	TNRB 230	TNRB 270	TNRB W110	TNRB W122	TNRB W210	TNRB W310
12:00 - 12:40	1-9	2-7	3-7	3-15	4-12	4-5	3-13	2-12	4-7	1-13	3-11
12:45 - 1:25	4-3	1-4	2-10	4-4	1-12	2-6	4-6	3-9	3-8	3-6	1-6
1:30 - 2:10	2-9	4-14	4-1	1-3	3-5	3-1	2-11	4-11	1-15	2-14	2-2

2:15 - 2:30	Break										
2:35-3:15	3-2	1-8	1-2	3-4	1-14	4-2	1-11	1-10	2-3	2-5	4-13
3:20-4:00	2-4	4-10	1-5	2-1	2-15	1-1	2-8	2-13	1-7	4-8	3-3
4:05-4:45	3-10	3-12	3-14	4-9							

Slack

A Slack workspace has been set up for you to communicate with instructors, TAs, the clients, and other students (see below). Please use this link:

https://join.slack.com/t/intex-winter-2023/shared_invite/zt-1srs8aycl-byuzgdCqERph3NlsnahTtA

You will be able to create your own channels for your team within this workspace.

Questions

Q & A

You are not able to ask Faculty and TAs for help with technical problems, but you are welcome to ask questions in regard to the case itself or the requirements. However, due to the number of groups and students involved, please post any questions on the "discussions_and_questions" channel in our INTEX2 workspace on Slack so that everybody can benefit from the response and that the faculty are not answering the same question over and over again to different groups. The clients may also answer questions here.

You may ask questions to other students working on the project (use the "discussions_and_questions" channel in our INTEX2 workspace), but you may not ask questions of former students who have completed a similar project in a past INTEX. Seeking help from previous students in any way or using any of their material would be considered cheating. You may ask questions of people in industry or other outside help, but the work must be your own.

Professor Wells will serve as the point of contact for administrative questions like "I don't see how to submit INTEX II, can you tell us?" but individual instructors will answer class-specific requirement questions.

INTEX Results & Awards

We will announce the winning team(s) next week.