Group Meeting Log **Project Group (Name/Number): The-GitGirls-Collective-7 - Group 7**

Week:	Date:	Activity/Task:	Time spent on Activity:	Team Members Involved (if applicable)	Completed?	Outcomes
1	11/13/2023	Group met during and beyond session time to plan project. Discussed SWOT for all members - Google Sheet completed with notes Grouped suggested topics for focussing dataset search Discussed what elements we need to look for in a good dataset Members did 30mins individual research on datasets, came back to present initial ideas on topics	90	All 6	Yes	Decided to spend next 2 days individually searching for datasets, and present any interesting ideas to group after class.
1	11/14/2023	Short meeting with full group for SK to run through possible use of ICO incident trends dataset, exploring positives and limitations (no API and basic visualisations already done)	45	All 6	Yes	Individuals to continue looking for datasets, posting links in Google Sheet to the best ones, highlight any really interesting ones on Slack.
1	11/17/2023	Met to pitch our best datasets from our solo research. SHS brought NHS API, SK brought NVD & CWE APIs, AMG the UN data resource and faostat resource (both very large datasets with APIs), NP brought two crime and justice datasets from ONS. Group agreed most interesting source at present was cybercrime / online fraud ONS dataset, but not sure if there was API interaction which we know is essential criteria. NP discovered during meeting there is the ability to query ONS data via API & volunteered to try to get that working over the weekend. Meanwhile other group members will look into the APIs of other good datasets which may be able to be intersected with the fraud database. We also discussed the marks weighting and that Project Documentation needs an an 'allowance' of 2 people and that it will be a good idea to keep on top of Documentation as we go along, TN volunteered to take charge and set up GitHub.	90	TN, SK, SHS, AMG, NP	Yes	NP to work on extracting the good dataset via API from ONS. Other members to see if their sources / best datasets can be extracted via API. TN to set up new GitHub repo and review documentation.
2	11/23/2023	Plan: short meeting to catch up on progress. Discussed everyones' best datasets at this point and timeline needing us to make an API dataset choice asap. Ran through Teleport offerings and limitations. Decided to code the API requests.	45	TN, SK, SHS, NP	Yes	SK to work on extracting data from the countries API. Other members think of potential research questions.
2	11/24/2023	Full group meeting to make decision regarding dataset. Presented code progress (.py files for API requests, SQL table, and JSON files resulting from first API requests). Looked at HW2 together, divided up starter tasks for that. Also divided up tasks for data-handling phases	90	TN, SK, SHS, AMG, NP	Yes	https://autumncfgdegree2023.slack.com/archives/C06507J102K/p1700858587915659
2	11/25/2023	Planned for short meeting to relay progress on initial tasks and tackle any problems / any interrelation needed between the tasks. A lot of work had been done so we shared progress on API code (SK), dataframe conversion code (NP), progress on HW2 and gant chart (AMG), high level SQL database planning and EER diagram (TN) and SQL table creation / deeper adjacent dataset investigation (SHS). We discussed challenges (JSON > dataframe) and brainstormed way to streamline work (going from csv where possible straight to pandas/jupyter rather than JSON > SQL > jupyter). We also discussed the	105	TN, SK, SHS, AMG, NP	Yes	
2	11/26/2023	Meeting discuss NP progress regarding dataframe, clarification of requirements (whether or not SQL needed or useful), worked jointly on question formulation as we could now see our dataset in CSV (being specific), looked at Data Questions tab in Google Sheet, discovered needed to convert currency of API returned salaries to GBP	75	TN, SK, SHS, AMG, NP	Yes	SK perform currency conversion on salaries TN to construct SQL database if appropriate for data
3	11/28/2023	Meeting to discuss progress on code and HW2 and SQL database	75	TN, SK, SHS, AMG, NP	Yes	
3	11/29/2023	Meeting about missing values from the API which was used for the SQL Database. Nicola created a presentation to explain what happened and to ask for the group confirmation before updating the API code.	45	TN, SHS, AMG, NP	Yes	Tatiana to finish the SQL database. Samantha to submit homework 2.
4	12/4/2023	Meeting to distribute the work going forward. SHS began typing all members contributions for creating a 'roadmap' of tasks t	75	TN, SHS, AMG, NP	Yes	Have another meeting tomorrow with SK to assign tasks to team members.
4	12/5/2023	Meeting with group to allocate cleaning, analysis, visualisation tasks. SK caught up in on progress of group over weekend on SQL database, location of updates to files in GitHub and group's creation of document to solidify analysis questions (Google Doc "Project Plan"). NP made presentation as had uncovered some serious outliers in our data which would skew our analysis if left unhandled, and presented a solution appropriate for the time we have left for the project. Group agreed ideally we'd go back and do more exploratory analysis on our data to see if there was a more thorough way to relativise or clean the data, however in the 4th week we agreed as a group we had to move forward with what we have, even if this means streamlining the subset of data we use or tailoring our questions relative to what the data can give us answers to. Also collectively decided on specific list of IT jobs to focus on, and how to streamline our data to the most relevant instances (i.e. classifying our countries as 'proper countries' vs 'regions' which may have been included). Each of us have a cleaning / visualising / analysis or documentation task to do and present back on Friday evening.	75	TN, SK, SHS, AMG, NP	Yes	TN to start on Project Documentation SK to work on heatmap of missing values SHS to check classification of countries in our list, add column to SQL Countries table, so that we can exclude any 'regions' or odd territories which may skew results. NP to continue work on code that produces average salary for each country with with to relativise individual interesting salaries. Looking for any further issues, or patterns. Look at ML analysis only if there is time. AMG to work on box plot for outliers and histogram for data

4	12/8/2023	Group meeting to share analysis progress, visualisations created, concerns etc. Confirmed exactly which version of the dataset files we need to be using. Discussed if we were going to be able to / need to incorporate the WorldData cost_of_living data. We brought together observations about how we were feeling about the quality of the dataset (boxplots and histograms created by people were shared) and how we may need to adapt the questions we ask. NP came up with some very clever suggestions for how we can usefully make sense of the data we have despite the apparent variablility across countries for the local currency salary figures (internal consistency thankfully was observed from the visualisations).	80	SK, NP, AMG, SHS	Yes	SK to finish and upload Juptyer Notebook version to GitHub for others to add their contributions to. NP to work on idea for analysis on a country-country 'internal' basis, which can then be compared to others.
4	12/9/2023	Group meeting to distribute documentation tasks, final visualisations and analysis tasks. SHS gave feedback on testing the Jupyter notebook which highlighted the need for it to be tested by all team members, and also prompted a run through of the notebook, updating filepaths from code that had come from different team members. Also discussed how best to organise the Juptyer Notebook file dependencies within GitHub.	50	TN, SK, SHS, AMG, NP	Yes	AMG to upload charts to Juptyer AMG and SHS to test notebook, provide feedback on commentary and comments, add in any extra analysis or comments, reorganise sections etc TN to start on documentation and be in charge of divvying up different sections. TN to complete section for Jupyter notebook regarding SQL database construction. NP as per previous discussion, just to focus on visualisations. SK to support NP with coding, testing, any other delegated tasks. SK to look at potentially altering NP code in Notebook to take a list of jobs as arguments rather than different function per job title.
5	12/11/2023	Meeting to run through documentation progress, new visualisations (ML, yay!), distribute remaining tasks, revisited intial question focus again as a group to make sure our visualiastions and analysis were relevant. Ran through the marks scheme breakdown and proportions. Discussed how our understanding of the dataset has evolved during the project, it's limitations and how we have adapted our analysis to the dataset rather than forcing questions that we were interested in but possibly don't have answers from our data. Ended with distributing remaining things to do, agreed an aim of trying to get the project finalised by Friday to give us all time to revise for the exam.	80	TN, SK, SHS, AMG, NP	Yes	
5	12/13/2023	Met to go over project requirements and allocate last pending tasks	70	TN, SHS, AMG, NP	Yes	
5	12/14/2023	Discussed achieving specific documentation requirements	15	NP, TN, SHS	Yes	
5	12/16/2023	Discussed final tasks for project submission - documentation tasks, Jupyter Notebook final details including last analyses, attatching powerpoints rather than including them in the doc.	60	TN, SHS, AMG, NP	Yes	Delegated and clarified tasks
		Total Meeting time (hrs): 19.42				

TEAM MEMBERS & THE ROLES IN DETAILS
Sam Kerr - In a hands-on coding role, I assumed primary responsibility for a Python API program, adapting it to evolving data project needs, such as incorporating currency conversions, and laying the groundwork for initial DataFrames. I seamlessly transitioned the program into a Jupyter Notebook (Part 1) for enhanced flexibility. Collaborating closely with Nicola P, we engaged in reciprocal code review and testing, integrating mutual improvements into central files. Early in the project, I compiled project requirements into a task list and diligently maintained a meeting summary log. Alongside team members, I invested significant effort in the search for a high-quality dataset. Beyond coding, my main focuses encompassed organisational aspects and strategic planning, culminating in the production of visualisations and analysis to contribute to the project's overall success.
Nicola Prevost - I streamlined data processing by integrating an API into the DataFrame, addressing compatibility issues with outdated ISO_alpha2 codes in Namibia for the currency converter. This paved the way for robust data analysis, including currency conversions and rectifying discrepancies. The cleaned dataset then fueled dynamic visualizations—bar charts, stacked bar charts, and scatterplots—to distill complex insights. Additionally, I delved into machine learning, implementing a Support Vector Machine (SVM) model for predictive analytics. This dual approach showcases my end-to-end proficiency in data science, from efficient data handling to advanced analytical
Tatiana Ngamba - I took on the lead role in data cleaning and normalization, handling extensive datasets and utilising Excel for data cleansing. Additionally, I created an ER diagram to design the SQL database structure, subsequently creating and implementing the SQL database which has been used for our data visualization. My responsibilities extended to project documentation, where I lead the creation of a comprehensive document by ensuring all sections were populated with relevant information. Furthermore, I played a pivotal role in project management, overseeing the group's progress, advocating for regular meetings, and asking relevant questions to drive the project forward.
Alicia Monge Grasa - I conducted thorough data cleaning and analysis using Jupyter Notebook, ensuring data integrity and uncovering meaningful insights through further data analysis and various visualizations. Simultaneously, I took charge of project management, overseeing timelines and coordinating team efforts to maintain project momentum. This integrated approach highlights my proficiency in hands-on data tasks and effective project leadership, ensuring a comprehensive and well-managed project lifecycle.
Samatha Hughes-Stanley - I authored a Jupyter notebook for our Homework 2, using markdown and coding to integrate formatted images, enhancing our documentation's clarity and visual appeal. Additionally, I developed a SQL database for our Country Codes table, vital for including population density data in our analysis. I also tackled the challenge of updating our gender pay parity data, where I curated a new, precise CSV file for ease of integration into our analysis, rectifying discrepancies and outdated information from our previous data sources. In terms of team collaboration, I contributed ideas, concerns, datasets, files via GitHub, write-ups and supported team members including troubleshooting errors.

Name: Tatiana Ngamba **Project Group (Name/Number): The-GitGirls-Collective-7 - Group 7**

Week:	Date:	Activity/Task:	Time spent on Activity:	Team Members Involved (if applicable)	Completed?	Notes:
1	*11/13/2023*	*Created a zoom meeting and povide the link to the group.*	*15*	*Tatiana*	*Yes*	Completed
1	*11/13/2023*	*Individually research data for Cybersecurity. And explained my findings to the team via zoom.*	30	*Tatiana*	*Not completed*	*Still deciding and it was decided we can do more research and have a meeting on 14/11/2013. After the meeting on Tuesday, I suggested that we take wednesday and thursday to reserch our topic further. and we can meet up at 5:30pm on friday to choose a direction we ould prefer to take.*
	14/11/2023 & 15/11/2023	*Whilst researching Cybersecurity data sets and API further. I reaslised that my knowledge for this field was not great and I could't fully understand what I was looking at. I decide to search for different data sets to see it there would be a better topic that we could explore further. I search for e-commerce topics by reading articles on vogue and Business of Fashion. I did come accross a subject that I thought was interesting. This was called 'E-commerce Fraud' - friendly Fraud for retail brands. However, whilst looking for data and API for this topic, there was not much data avalaible. I did come accross Credit cards detection fraud but the data sets was not suficientl for this project.*	180	*Tatiana*	*Not completed*	*We are still deciding which direction to take.*
1	*11/16/2023*	I downloaded and updated the project activity log spreadsheet. I added six tabs for each member and wrote all of our names and project name/number. I saved and uploaded the final version to our Slack group for easy accessibility.	30	Tatiana	Yes	Completed
1	*11/18/2023*	I created a new GitHub Repo. I added all group members, added our marker & CFG	30	Tatiana	Yes	Completed
1	*11/18/2023*	I create two projects managers (Project & Homework two) and added the breakdown tasks.	90	Tatiana	Yes	Completed
1	*11/18/2023*	Started the project Documentation	45	Tatiana	*Not completed*	Need to start documenting each section correctly.
1	*18/11/2023 & 19/11/2023*	I have spent the whole night/ Morning searching for relevant data set and API for our project/ chosen topics and I have been unable to find any suitable. I have even tried to access details via GET request on python and there wasn't much luck. So I decided to research more on a background that I have more knowledge in, which is Education and fashion. Education - unfortunately, AQA does not display their data sets about students & there is no way of getting any API. I was going to ask my friend to share some data as she is still working there but I did realised that this would be a data breach to be honest. hence why they would never share historical and current data set. Fashion - In regards to Fashion, I started reading more articles and i came across Instagram API, which can be use to see people's basic post information (https://developers.facebook.com/docs/instagram/) which use to analysing information iff we wanted. I came across H&M API - after further reading, they will only grant access to developers if it would bring them customers. Then I came across Asos API and I was able to find some data set via Kaggle. API - https://rapidapi.com/apidojo/api/asos2/ Data sets - https://www.kaggle.com/datasets/trainingdatapro/asos-e-commerce-dataset-30845-products Data Sets - https://data.world/tatiana-e93/project-cfg/workspace/file?agentid=opensnippets&datasetid=asks-products-dataset&filename=sample_asos_data_jun_2021.csv	300	Tatiana	Not completed	We will need to meet up as a group and finalise our lastest findings
2	11/21/2023	Searched for accessible API for the group project and found several free API.	60	Tatiana	Completed	We selected one of the public API for our project
2	11/25/2023	My focus was to create an SQL Database for our project using all the data from the API source and other sources. I used a spreasheet as a way to normalised the data and created and ER diagram which was shared to the team for approval before writing the SQL code.	180	Tatiana	Not completed	We decided to reduce the number of tables. The original ER diagram had 10 tables, which wouldn't have been feasible for our data visualization. After showing this to the group, it was decided that we reduce the tables and add more data for certain tables.
2	11/26/2023	I did a few research on how we could phrase our main concept question for question 1—homework 2. I created a Jupyter notebook for our homework and added all the homework questions and the relevant answers that was done by the group. This documnet to shared in our slack channel, so that Nicola and Samantha could updated it with their answers.	180	Tatiana	Not completed	We still need to decie on our concept questions. Question 3 and 4 and nearly done but we still need to add more to question 1 and 2.

3	11/27/2023	Redid the ER diagram and sent the updated version to the group channel for approval and final confirmation from the group.	60	Tatiana	Completed	After sharing this, Sam and Alicia had some suggestions on how to create a primary key for the job table. as the job id and the job titles are the same things.
3	11/28/2023	I examined the data and renormalized the final version of our SQL table. I moved columns across to ensure that we only had 3 tables and less data duplication. I struggled with the Gender pay column as I need that column to match the data that was already for the countries table and we had this data for some countries but not all. I went through each column of the data which took a very longtime. After class, we had a team meeting and Alicia showed me how to do a VLOOK UP . and I carried on working on these spreasheets ads the final data will be inserted to our 3 SQL data.	240	Tatiana & Alicia	Not completed	finished Countries spreadsheet but now had to create the job and the salaries table.
3	11/29/2023	Whilst analysing the data and adding the relevant columns to each table spreadhseet, I realised that we were missing data salaries table and flagged this to the team.	120	Tatiana	Not completed	
3	11/30/2023	NIcola provided an updated spreadsheet with all the data values including the missing values which was then used to update the spreadsheets that I was working on.	180	Tatiana		
3	12/1/2023	Wrote the SQL Code for our 3 tables and inserted values into these three table. I exported each tables as a CSV files and uploaded this on github.	360	Tatiana	Completed	
3	12/1/2023	Reviewed Homework 2 and added my comments to the group channel as Samatha was going to submit this homework.	60	Tatiana	Completed	
3	12/3/2023	I was going to merge all the SQL tables together via excel and save it as a CSV but turns out I am unable to do that because there are multiple sheets. Whilst doing this, I realised that SQL only exports 1000 rows of data and we have more than 10,000 rows of data per table for Jobs and salaries. so I had to export multiple spreadsheets and saved it accordingly in one big spreadsheet. Now we have all the data in each spreadsheet (Countries, Jobs and Salaries). I have uploaded these spreadsheets in the 'SQL branch' as well.	180	Tatiana	Completed	
4	12/5/2023	Reading all the individual spreadhseets on Jupyter notebook ensure that it works and is ready for us to use.	60	Tatiana	Completed	
4	12/9/2023	Created the project documentation and all all the relevant bullet point. I created the project title, created a picture logo for the documentation, and organised the documentation. I added the introduction, aims and objectives.	240	Tatiana	Not completed	There is still a lot that will need to be done for the documentation.
4	12/10/2023	Wrote down the summary of how the SQL table was structured including how the data was processed and the missing values. Added the SQL to the GLOBAL IT SALARY ANALYSIS (Git-Girls-Collective-7) - Notebook Part 2 (Data Analysis)	240	Tatiana	Completed	
5	12/11/2023	Working on Project documentation	90	Tatiana	Not completed	
5	12/12/2023	Working on Project documentation and created the outline for the main branch READ ME file on Github to outline where all the relevant will be displayed.	210	Tatiana	Not completed	
5	12/13/2023	Finding the correct colours for our stacked bar chart. Testing and running our Jupyter notebook. Adding suggestions and team meetings.	220	Tatiana	Completed	
5	12/15/2023	Recreated a new project documention to ensure that we meet the documention criteria	12 Hours	Tatiana	Not completed	
5	12/16/2023	Focused and worked on documentation whilst consulting the group on certain sections.	7 hours	Tatiana	Not completed	
5	12/17/2023	Finished the reporting section of the documentation	2 hours	Tatiana	Completed	

Individual Time **Total time: 78 Hours**

: Sam Kerr **Project Group (Name/Number): The-GitGirls-Collective-7 - Group 7**

This activity log is for you to document your individual contributions and ideas you have done towards your final group project. This will be submitted to your instructors along with your project code. You can document how many minutes you spent working on your code, when you had team meetings outside of the class sessions, or use this as a way to take notes for your own viewing. This is to be completed **individually.**

Week:	Date:	Activity/Task:	Time spent on Activity:	Team Members Involved (if applicable)	Completed?	Notes:
1	11/13/2023	Group met during and beyond session time to plan project. Discussed SWOT for all members - Google Sheet completed with notes Grouped suggested topics for focussing dataset search Discussed what elements we need to look for in a good dataset Members did 30mins individual research on datasets, came back to present initial ideas on topic Decided to spend next 2 days individually searching for datasets, and present any interesting ideas to group after class	90	All 6		
1	11/13/2023	Spent 1.5hrs individually looking into crypto sources and latterly cybersecurity data sources. Found one potentially interesting dataset and I was researching ways to interact with it (MitreAtt&ck API and Python library)	90	SK		
1	11/13/2023	Emailed address listed on the gov.uk Cyber security breaches survey 2023 page to ask if a dataset was available	10	SK		
1	11/14/2023	Adding links to Google Sheet Braindump Dataset Links tab and set up Activity Log Individual	20	SK		
1	11/14/2023	Initial explore of ICO Incident dataset, explored idea of mapping to MitreAtt&ck framework. Researched STIX and MitreAtt&ck. Emailed ICO to see if there are any more granular versions of the dataset available for analysis	60	SK		
1	11/14/2023	Meeting with full group to run through possible use of ICO incident trends dataset, exploring positives and limitations	45	SK		
1	11/15/2023	Research into NVD CVE database and API functionality. Wrote post to group on Slack	60	SK		
1	11/16/2023	Research into CWE and links between CWE and CVE database. Info posted to Slack. Messaged Programmes Team regarding Conclusion section of Project Documentation. Wrote to assessor to clarify & received answer	60	SK		
1	11/17/2023	Put HW2 into Google Sheet, broken down. Added tab to Google sheets with thoughts about meeting agenda for tonight	45	SK		
1	11/17/2023	Wrote Python script to make API call to NVD CVE, tested. Both a sample size and a larger API call for full DB. Made script to calculate filesize and time estimates, to inform how to batch data. Started planning re modifying API call to move through batches of 120 days, as a flexible way of scoping our dataset (stopped as group may not decide to use this dataset). Started researching how to parse JSON file data into SQL, i.e. nested dictionary structures (explored csv format initially, not a good idea, defo best to go straight to SQL). Stopped as may not be using this dataset	240	SK		
1	11/17/2023	Meeting with full group (see Meeting tab for full details)	90	5		
1	11/18/2023	Set up meeting tab in Project Activity Log. Slack post to group re GitHub branch organisation and work over weekend.	30	SK		
2	11/22/2023	Spent time looking at dataset links shared by team members on Slack. Found Telaport, spent time reading the API documentation and looking around the JSON objects seeing what information was available via API vs their website. Looks promising, particularly salary data for countries and 'urban area' metrics. Could defintiely find some interesting questions to ask. Python code might be laborious due to multiple categorisation levels. But not starting work on coding API request unless group are interested. Made a Slack post with links. Quick comms with group arranging next meeting.	90	SK		
2	11/23/2023	Group met to make a decision regarding everyone's best datasources at this point, we decided to use Teleport's data as the basis and try to find smaller country-specific statistics to cross reference. Agreed I would start work on the API whilst other group members looked for relevant intersecting data / thinking about questions to ask	45	4		Do API request code
2	11/24/2023	Wrote Python script to make API calls to Teleport, grabbed the list of countries, the overview country statistics and the salary data for each country. Also created git folder and sync'd with remote repository, uploaded data and .py files.	240	SK		
2	11/24/2023	Group meeting to share progress, allocate tasks for HW2 and starting on the data handling. I will be starting on HW2 Q4 and finishing the API call code.	90	All 5		Finish API request HW2 Q4

2	11/25/2023	Updated Dataset Links in Google Sheet to separate out discarded sources and additional sources now relevant to the main Teleport Dataset. Added links shared in Slack to this Created Data Questions sheet in Google Sheet. Brainstormed 4 questions in a lot of detail (Q, datasets used, analysis steps, visualisation, purpose).	90	SK		
2	11/25/2023	Rewrote .py code for teleport_API_requests_countries to improve consistency, including writing a new helper function, renaming log files etc. Created new GitHub branch SamK_Proj_Code and uploaded work there, created Pull Request	120	SK		
2	11/25/2023	Worked on urban_areas py code (later abandoned, not going to be useful)	90	SK		
2	11/25/2023	Meeting to check in on progress on API code, SQL database design, additional data sources. Dropped requirement for urban_areas extra data to focus on the salaries data. Very small amount of help given to NP (just explaining current function version) while she was working on upgrading API request code to go straight to DF vs JSON as my version was set up to do.	105	All 5		SK do / finish HW2 Q4
2	11/26/2023	Wrote HW2 Q4	90	SK		
2	11/26/2023	Meeting with group to discuss NP progress regarding dataframe, clarification of requirements (whether or not SQL needed or useful), worked jointly on question formulation (being specific), looked at Data Questions tab in Google Sheet, discovered needed to convert currency of API returned salaries to GBP, I took that on	75	All 5		
2	11/26/2023	Worked on currency conversion, involved overhaul of .py file. Had to add back in a function that was discarded, update NP API > DF functoin to include currency code (relying upon a different json file and function), write whole new function to use new json file of currency conversions to GBP to interact with the csv file, to output 3 new columns displaying salaries converted to GBP. Cleaned the unnecessary log and py-generated files out of github branch, created pull request with explanatory comments about the changes.	130	SK		
3	11/27/2023	Comms with group members about code and SQL database	20	SK, TN, NP		
3	11/27/2023	Wrote a new function api_currency_rates to bring the creation of the currency_rate.json inside the python code and alleviate the need for an external file downloaded seperately. Also now makes the code completely dynamic for creating a 'fresh' dataset each time. Made sure GitHub branches were uptodate & pushed updated code. Created new folder for datasets	60	SK		
3	11/28/2023	Spent about 3hrs starting work in Jupyter NB trying to join our datasets into a large dataframe with all information which can be refined to be analysed. SQL integration code (nightmare), some basic cleaning and discarding of data, backup output csv's saved at all stages. Uploaded to separate folder on my branch in github. Started this for my own practice and as I'll be away this weekend so needed to make a contribution early.	180	SK		
3	11/28/2023	Meeting with group to discuss finalising HW2 and next steps for project	75	All 5		
3	11/29/2023	5 hours overhauling python code (wrote main.py, organised other code into utils). Updated functions to produce two timestamped (so unique) files to better control versioning. Recommented code. Continued working on Jupyter Notebook on loading, joining and cleaning stages. Worked on commenting Notebook as I would not be around to explain the code this weekend. Joined country, salary, gender pay and cost of living datasets into dataframes. Produced csvs to convert into MySQL tables. Updated a couple of group members with more detail about what has been done, and put a message on Slack group, and uploaded files to GitHub branch so that everyone can access. This hopefully puts group in the position where they can analyse all our 4 datasources together and start asking questions / making visualisations.	300	SK		Had to get ahead on my contribution this week as I am away over weekend.
4	12/5/2023	Spoke with NP and looked at updates to Python code for cleaning troublesome values. Closed 2 pull requests. Merged NP code changes into the most uptodate Python code (from my branch) and created pull request to merge into the Project_Code branch	60	SK, NP		
4	12/5/2023	Meeting with group to allocate cleaning, analysis, visualisation tasks. SK caught up in on progress over weekend on SQL database, location of updates to files in GitHub and groups progres on document to solidify analysis questions (Google Doc "Project Plan"). NP made presentation as had uncovered some serious outliers in our data which would skew our analysis if left unhandled, and presented a solution appropriate for the time we have left for the project. Group agreed ideally we'd go back and do more exploratory analysis on our data to see if there was a more thorough way to relativise or clean the data, however in the 4th week we agreed as a group we had to move forward with what we have, even if this means streamlining the subset of data we use or tailoring our questions to what the data can give us answers to. Also collectively decided on specific list of IT jobs to focus on, and how to streamline our data to the most relevant (i.e. classifying our countries as 'proper countries' vs regions which may have been included). Each of us had a cleaning / visualising / analysis or documentation task to do and present back on Friday evening.	75	All 5		
4	12/5/2023	Updated activity log and meeting tab with summary	15	SK		

4	12/8/2023	Spent all day refining existing sections of Juptyer Notebook draft, reorganised it into sections, recommented, took out discarded or outdated parts. Took quite a long time to decipher the path the data had taken through the intitial dataframes created by combining the API call data and various datasets (which had errors) through to the SQL Database and the final xlxs output from those tables. Added in new section referencing SQL Database, and recreated DataFrames from the SQL xlsx file. Completed basic analysis, some checking and cleaning. Created a Heatmap to show the missing values and wrote a small analysis section. Spent time during the day working with NP to add in her code and findings from her analysis. Spent about 1.5hours interrogating the data, going back to the original API requests, the currency conversion, all versions of the dataframes etc to work out if the extreme values which seemed to be undermining the dataset were the result of error or if they were legitimate. Spent time sampling the extreme low datapoints (iran, vietnam, laos) and researched what these countries average annual salaries are, what their national minimum wage is in their own currencies. Compared this data to the API data. Also investigated how much of our dataset might be affected based on the 'dollar a day' heuristic. From this research I wrote an email to Teleport asking for help. Thought about the different options we could take from here to discuss at the group meeting.	420	< and that's conservative! SK (NP)		
4	12/8/2023	Sent email to Teleport asking for insight about their API data	25	SK		
4	12/8/2023	Group meeting to share analysis progress, visualisations created, concerns etc. Confirmed exactly which version of the dataset files we need to be using. Discussed if we were going to be able to / need to incorporate the WorldData cost_of_living data. We brought together observations about how we were feeling about the quality of the dataset (boxplots and histograms created by people were shared) and how we may need to adapt the questions we ask. NP came up with some very clever suggestions for how we can usefully make sense of the data we have despite the apparent variablility across countries for the local currency salary figures (internal consistency thankfully was observed from the visualisations).	80	SK, NP, AMG, SHS		SK to finish and upload Juptyer Notebook version to GitHub for others to add their contributions to. NP to work on idea for analysis on a country-country 'internal' basis, which can then be compared to others.
4	12/8/2023	Completing personal log and meeting log	10	SK		
4	12/9/2023	Group meeting to distribute documentation tasks, final visualisations and analysis tasks. SHS gave feedback on testing the Jupyter notebook which highlighted the need for it to be tested by all team members, and also prompted a run through of the notebook, updating filepaths from code that had come from different team members. Also discussed how best to organise the Juptyer Notebook file dependencies within GitHub.	50	TN, SK, SHS, AMG, NP		AMG to upload charts to Juptyer AMG and SHS to test notebook, provide feedback on commentary and comments, add in any extra analysis or comments, reorganise sections etc TN to start on documentation and be in charge of divvying up different sections. TN to complete section for Jupyter notebook regarding SQL database construction. NP as per previous discussion, just to focus on visualisations. SK to support NP with coding, testing, any other delegated tasks. SK to look at potentially altering NP code in Notebook to take a list of jobs as arguments rather than different function per job title.
4	12/9/2023	Updated Jupyter notebook with group changes from meeting and slight changes to GitHub branch organisation as per group decisions	15	SK		
4	12/9/2023	Attempted to assist with troubleshooting SQL cell via Slack	15	SK		
4	12/9/2023	Reformatted NP def analyst() and def engineer() functions to take in a list of the relevant it roles and produce a chart for each, side by side.	90	SK		
4	12/10/2023	Slack comms organising the structure of the notebook	30	SK		
4	12/10/2023	Re organised Juptyer Notebook to include new version of NP code. Pushed to github, messaged on slack explaining changes	90	SK		
4	12/10/2023	Created new Juptyer Notebook for Python API code. Full reorganise of code, splitting out all commentary, docstrings etc. Did several full tests and added in commentary just for the notebook and some additional print statements which help the notebook format. Direct push upload to GitHub Project_Code branch.	120	SK		
4	45270	Wrote the Data Collection section for the Project Documentation, detailing sources and methods. Some gaps to be completed by other team members and formatting / length was not finalised. Emailed Documentation Lead some resources for other sections	30	SK		
5	12/11/2023	Revisited task breakdown generated at the beginning of project to ensure we are aware of all the requirements. Did some final checks regarding data integrity, revisiting the initial expectations and assumptions about the dataset from Teleport.	30	SK		

5	12/11/2023	Meeting to run through documentation progress, new visualisations (ML, yay!), distribute remaining tasks, revisited intial question focus again as a group to make sure our visualiastions and analysis were relevant. Ran through the marks scheme breakdown and proportions. Discussed how our understanding of the dataset has evolved during the project, it's limitations and how we have adapted our analysis to the dataset rather than forcing questions that we were interested in but possibly don't have answers from our data. Ended with distributing remaining things to do, agreed an aim of trying to get the project finalised by Friday to give us all time to revise for the exam.	80	TN, SK, SHS, AMG, NP		
5	12/12/2023	Wrote the Background section of the Project Documentation. Not finished, likely to need tidying by Project Documentation lead.	45	SK		
5	12/12/2023	Comms with team on Slack catching up on last nights' meeting	30	SK		
5	12/15/2023	Comms with team on Slack organising final task list. I will be in charge of final ipynb tidy-up	30	SK		
5	12/15/2023	Another 5 hours spent testing both Jupyer Notebook 1 (API code) and Notebook 2. Fully tested the API Notebook again from start to finish, tweaked some of the file storage locations in the code to make it more organised. Overhaulled notebook 2, went through from start to finish added to commentary, reorganised sections, tidied formatting, fixed errors (references to dataframes), put in NP heatmap code. Did all the section numbering, Changed the Section 6 headings to questions. Fully tested and working. Added notes for the README along the way. Deleted duplicate chart and any unneeded comments. Added in some comments where some further organisation or analysis is needed. Also tidied the Project_Code branch on GitHub, making sure all the correct files were in data folder, cleaning out unnecessary ones, tidying file names, making sure it all matched with the Notebook code. I am DONE!!	300	SK		
5	12/16/2023	Comms with team on Slack	15	SK		
5	12/17/2023	Reviewed README file and added a few small tweaks and comments, created new temp branch and created pull request with a summary of changes document supplied on Slack for speed of review. Reviewed the others' amazing work on Canva Project Documentation over the past few days. Caught up on Slack messages on past few days, short discussion re presentation timings. Tidied Requirements Google Sheet Task list and finished my activity log	60	SK		

Total time (hrs): 72.58333333

Individual time (hrs, minus meeting hours)
Attended 15hrs meetings of 19hrs total: **57.50**

Name: Samantha Hughes-Stanley **Project Group (Name/Number): The-GitGirls-Collective-7 - Group 7**

This activity log is for you to document your individual contributions and ideas you have done towards your final group project. This will be submitted to your instructors along with your project code. You can document how many minutes you spent working on your code, when you had team meetings outside of the class sessions, or use this as a way to take notes for your own viewing. This is to be completed **individually.**

Week:	Date:	Activity/Task:	Time spent on Activity:	Team Members Involved (if applicable)	Completed?	Notes:
1	11/13/2023	Met with my group to discuss initial ideas. We discussed shared interests, we like Cyber Security, Pharmaceuticals and CryptoCurrency. We left the call to each do individual research into our chosen topics. We recovenened to discuss our findings. We had issues with pay walls, quality of data, age of data, etc. We agreed to meet again on Tue to discuss any futher findings.	120	All	Yes	
1	11/14/2023	Met with the group again, Sam had found a good Cyber Security data set but we didn't know yet how to incorporate an API and meet all the criteria with it. Booked another meeting on Friday when we've all had some more time to find other possible data.	60	All	Yes	
1	11/17/2023	Spent time researching possible data sets for the project, ready to feedback at the meeting tonight.	180	Me	Yes	
1	11/17/2023	Met with the group to discuss all our findings and next steps	90	All	Yes	
1	11/18/2023	Researched more datasets, for adding on to the main Teleport data. Found UN data on birth rates and life expectancy. Looked at potential keys for joins	60	Me	Yes	
2	11/23/2023	Group met to make a decision regarding everyone's best datasources at this point, we decided to use Teleport's datasets	45	All	Yes	
2	11/23/2023	Set up country codes SQL database and table, shared this to Github	60	Me	Yes	
2	11/24/2023	Group meeting to share progress, allocate tasks for HW2 and starting on the data handling. I will be starting on HW2 Q4 and finishing the API call code.	90	All	Yes	
2	11/24/2023	Trialled importing a CSV to MySQL and creating a join on the two datasets, encountered some problems, troubleshooting solutions	90	Me	Yes	
2	11/25/2023	Meeting to check in on progress on API code, SQL database design and additional data sources. I shared the issues I encountered with the additional datasets and MySQL, we agreed on how we wanted to move forward this and I will lead on that task	105	All	Yes	
2	11/25/2023	Found better data on gender pay parity to overcome issues found on the other datasets, manually created a CSV, this involved reformatting data copied from the World Economic Forum website. Submitted to GitHub	45	Me	Yes	
2	11/25/2023	Wrote a first draft answer to Q3, utilising some notes Nicola had made and building on them to complete the criteria. Submitted to Slack	60	Me	Yes	
2	11/26/2023	Catch up with team to share progress and plan next steps	75	All	Yes	
2	11/26/2023	My task is to finish off Homework Question 1, Nicola will give her input and I will then add my input to complete the question	60	Me and Nicola	Yes	
3	11/28/2023	Attended meeting to discuss progress on code and HW2 and SQL database	75	TN, SK, SHS, AMG, NP	Yes	
3	11/29/2023	Attended meeting about missing values from the API, Nicola shared a presentation and the group agreed a way forward. I took the task of formatting HW2 into a Jupyter notebook and submitting it.	45	TN, SHS, AMG, NP	Yes	
3	11/29/2023	Shared a google doc with the team, with all homework contributions combined. I implimented the changes suggested by the team, for them to read and review final draft before formatting into Jupyter notebook	90	SHS	Yes	
3	12/30/2023	Created Jupyter notebook for HW2, inserting images for better understanding. Share final draft with the team to check it worked well for them when they downloaded the zip file as the assessor would. Discovered some images were showing well for me, for not for the group, so re-formatted them correctly, and submitted the final notebook via GitHub	160	SHS	Yes	
4	12/4/2023	Meeting to distribute the work going forward. I wrote up a 'roadmap' of tasks for the group, directed by the meeting's discussiong, on a google doc for us all to see and edit	75	TN, SHS, AMG, NP	Yes	
4	12/5/2023	Meeting with group. See meetings tab for further details	75	TN, SK, SHS, AMG, NP	Yes	
4	12/8/2023	Set up a new zoom link for the group meeting, while Tatiana wouldn't be present, as she usually hosts	5	SHS	Yes	
4	12/8/2023	Group meeting, see meetings tab for further details	80	SK, NP, AMA, SHS	Yes	

4	12/9/2023	Troubleshooting errors. I finally discovered that there was a version of Pandas which I had installed, which was incompatible with SQLAlchemy's OptionEngine. Updated Pandas and the original code then worked without errors. Informed the team and contributed a small write up to advise our assessor to be aware of this issue, and ensure they had the most current version of Pandas	180	SHS	Yes	
4	12/9/2023	Group meeting - see other tab for more detail. I updated the team regarding the errors I had and the solution I'd found. Also suggested new default file paths in the Notebook, since we'd changed how the folder was organised, this now runs with no	50	TN, SK, SHS, AMG, NP	Yes	
5	12/11/2023	Meeting, see meetings tab for more detail	80	TN, SK, SHS, AMG, NP	Yes	
5	12/13/2023	Meeting, see meetings tab for more detail	70	TN, SHS, AMG, NP	Yes	
5	12/14/2023	Meeting, see meetings tab for more detail, I agreed to take on some more documenation and analysis of our visualizations	10	TN, SHS, NP	Yes	
5	12/15/2023	Updated my log, caught up with latest progress of the team by reading through the newest Jupyter Notebook and documentation. Discussed creating some more visualizations with Tatiana.	90	SHS	Yes	
5	12/15/2023	Wrote analysis for 5 visualisations, attempted to add them to the Jupyter Notebook along with Nicola's heatmap, but encountered errors. Abandoned that version to avoid risking problematic code being present in the doc. Shared my visualisation analyses in a google doc via slack and let the group know I can give that another try later, or if they want to add these analyses in instead that's fine by me	210	SHS	Yes	
5	12/16/2023	Meeting, see meetings tab for more detail, my tasks will be to read through the documentation and write the conclusion	60	TN, SHS, AMG, NP	Yes	
5	12/16/2023	Writing the conclusion, drafting the results section, proof reading and editing the whole documentation	330	SHS	Yes	
5	12/17/2023	Reviewing all elements, catching up with Slack channel, preparing to submit project	90	SHS	Yes	

Total time: 48.58

Name: Alicia Monge Grasa **Project Group (Name/Number): The-GitGirls-Collective-7 - Group 7**

This activity log is for you to document your individual contributions and ideas you have done towards your final group project. This will be submitted to your instructors along with your project code. You can document how many minutes you spent working on your code, when you had team meetings outside of the class sessions, or use this as a way to take notes for your own viewing. This is to be completed **individually.**

Week:	Date:	Activity/Task:	Time spent on Activity:	Team Members Involved (if applicable)	Completed?	Notes:
1	11/13/2023	Met with my group to discuss initial ideas. We discussed shared interests, we like (	120	All	Yes	
1	11/14/2023	Met with the group again, Sam had found a good Cyber Security data set but we di	60	All	Yes	
1	11/17/2023	Researched and added 5 links to datasets on the group spreadsheet	120	Me	Yes	Still need to choose the final dataset to work on
1	11/17/2023	Met with the group to discuss all our findings and next steps	90	All	Yes	
1	11/18/2023	Tested calling API from OECD and UN datasets	120	Me	Yes	No succesful API calls
2	11/24/2023	Group meeting to share progress, allocate tasks for HW2 and starting on the data handling. I will be starting on HW2 Q4 and finishing the API call code.	90	All	Yes	
2	11/24/2023	Researched on additional dataset about wage gap	60	Me	Yes	OECD found but finally not used
2	11/25/2023	Designed, created and organized gantt chart	120	Me	Yes	
2	11/25/2023	Meeting to check in on progress on API code, SQL database design and additional data sources. I shared the issues I encountered with the additional datasets and MySQL, we agreed on how we wanted to move forward this and I will lead on that task	105	All	Yes	
2	11/25/2023	Ellaborated question 2 about target audience	90	Me	Yes	
2	11/26/2023	Catch up with team to share progress and plan next steps	75	All	Yes	
3	11/28/2023	Meeting with group to discuss finalising HW2 and next steps for project	75	All	Yes	
3	11/29/2023	Final version of target audience to HW2 doc	30	Me	Yes	
3	11/29/2023	Gannt chart into HW2 doc and description added	30	Me	Yes	
4	12/5/2023	Meeting with group to allocate cleaning, analysis, visualisation tasks. Outliers presentation + filtration for IT jobs	75	All 5	Yes	
4	12/7/2023	Histogram and box plots for IT salaries and gender pay, identify outliers in the dataset	180	Me	Yes	
4	12/8/2023	Group meeting to share analysis progress, visualisations created, concerns etc. Confirmed exactly which version of the dataset files we need to be using. Discussed if we were going to be able to / need to incorporate the WorldData cost_of_living data. We brought together observations about how we were feeling about the quality of the dataset (boxplots and histograms created by people were shared) and how we may need to adapt the questions we ask. NP came up with some very clever suggestions for how we can usefully make sense of the data we have despite the apparent variablility across countries for the local currency salary figures (internal consistency thankfully was observed from the visualisations).	80	4	Yes	
4	12/9/2023	Group meeting to distribute documentation tasks, final visualisations and analysis tasks. SHS gave feedback on testing the Jupyter notebook which highlighted the need for it to be tested by all team members, and also prompted a run through of the notebook, updating filepaths from code that had come from different team members. Also discussed how best to organise the Juptyer Notebook file dependencies within GitHub.	50	All 5	Yes	
4	12/10/2023	Visualization plots: box plots per country on salary and scattered plot combining pay parity and IT median salaries	180	Me	Yes	
5	12/11/2023	Meeting to run through documentation progress, new visualisations, distribute remaining tasks, revisited intial question focus again as a group to make sure our visualiastions and analysis were relevant. Ran through the marks scheme breakdown and proportions. Distributing remaining things to do.	80	All 5	Yes	
5	12/11/2023	Updates in the scattered plot by deleting outliers, add explanation and pull request into shared notebook in Github	120	Me	Yes	
5	12/13/2023	Update gantt chart and project log progress of past weeks	60	Me	Yes	
	12/13/2023	Met to go over project requirements and allocate last tasks	70	All 5	Yes	

5	*12/15/2023*	Included regression line to scattered plot, organized jupyter notebook per sections, added conclussions and plot interpretations, reran file. Started readme file	*240*	*Me*	*Yes*	
5	*12/16/2023*	Rerun code and completed some pending clarifications in notebook 2	*90*	*Me*	*Yes*	
5	*12/16/2023*	Discussed final tasks for project submission - documentation tasks, Jupyter Notebook final details including last analyses, attatching powerpoints rather than including them in the doc.	*60*	*4*	*Yes*	
5	*12/16/2023*	Write Readme file with clear instructions on how to run the project, re-review notebooks and project documentation and update project log and gantt chart	*360*	*Me*	*Yes*	
5	*12/17/2023*	Went over project documentation and final project log update	*60*	*Me*	*Yes*	

Total time: 48.17

Name: Nicola Prevost **Project Group (Name/Number): The-GitGirls-Collective-7 - Group 7**

This activity log is for you to document your individual contributions and ideas you have done towards your final group project. This will be submitted to your instructors along with your project code. You can document how many minutes you spent working on your code, when you had team meetings outside of the class sessions, or use this as a way to take notes for your own viewing. This is to be completed **individually.**

Week:	Date:	Activity/Task:	Time spent on Activity:	Team Members Involved (if applicable)	Completed?	Notes:
1	11/15/2023	Researched and added 3 links to datasets on the group spreadsheet	120	Nicola Prevost	Yes	Still need to select a specific area to hone in on
1	11/18/2023	Emailed ONS to request access to API	60	Nicola Prevost	Yes	Request was denied
2	11/20/2023	Researched more APIs and added one link to the group spreadsheet	60	Nicola Prevost	Yes	Limited resources for open API datasets
2	23/11/2023 - 24/11/2023	Researched more APIs, found 5 total and obtained data from 2 of them.	180	Nicola Prevost	Yes	Will discuss my findings at the group meeting tonight.
2	11/25/2023	Manged to get the API data straight into a DataFrame.	60	Nicola Prevost	Yes	
2	11/25/2023	Began Q3 of HW2 by writing notes with key points to consider.	60	Nicola Prevost (SHS)	No	Asked Sam HS to write my notes into sentences as she writes more eloquently than me
2	11/26/2023	I helped write up Q1 of HW2 and came up with a hypothesis for our project.	60	Nicola Prevost (SHS)	No	I will help Sam HS with HW2 by proof-reading and commenting suggestions.
3	11/29/2023	I wrote the summary for the group meeting.	10			
3	11/30/2023	I fixed the problem with data not pulling through correctly to the csv file.	200	Nicola Prevost	Yes	
3	11/30/2023	I created and gave a presentation on these proposed changes.	200	Nicola Prevost	Yes	
4	12/4/2023	I started a Jupyter Notebook where I experimented with the data and came up with a realistic plan for how to analyse the data.	120	Nicola Prevost	Yes	
4	12/4/2023	I wrote the summary for the group meeting.	10	Nicola Prevost	Yes	
4	12/5/2023	I began analysing the data and encountered problems with very low and very high salaries for different countries. I came up with a solution and created a presentation to give to the group about it.	220	Nicola Prevost	No	Completed on 06/12/2023
4	12/5/2023	Spoke to SK about merging my missing values fixes to the project_code branch.	25	Nicola Prevost (SK)	Yes	
4	12/6/2023	Following the presentation the evening before, I researched alternative ways to contextualise the very low and high values so that we could draw meaningful conclusions. I found that using the median instead of the mean of all salaries in the dataset for one country was the best way. I proceeded to write some code so that the countries for the IT jobs we had selected would be assigned one of three values; more than the median, same as the media and less than the median.	200	Nicola Prevost	Yes	
4	12/8/2023	Spoke to SK to explain my code and problems I had found during my initial analysis. Also had a meeting with SK to investigate which data we were all using as there were different versions circulating.	60	Nicola Prevost (SK)	Yes	
4	12/9/2023	I started the Machine Learning for our project, specifically simple linear regression and SVM with linear, polynomial and rbf kernels. I also looked into incorporating the pay parity data but it has no obvious relationship with the other variables in our dataset. Also, if we want to predict pay parity for other countries, we will have to assign new pay parity categories (low, medium, high) as SVM does not work with continuous data as the target variable.	300	Nicola Prevost	No	Completed 11/12/2023
4	12/10/2023	I brainstormed ideas for the analysis visualisation and started coding it.	200	Nicola Prevost	No	Completed 10/12/2023
4	12/10/2023	Coding a new idea for visualisations; a stacked bar chart for each IT job with the number of countries as the y variable and the colours as the more/same/less than median.	180	Nicola Prevost	Yes	
4	12/10/2023	Updated my activity log to reflect the last 3 weeks' work.	60	Nicola Prevost	Yes	
4	12/11/2023	I incorporated Machine Learning by generating an SVM model to predict levels of pay parity (high, medium and low) with continent and population density as the explanatory variables. I also generated a new scatterplot showing the relationship between continent, pay parity and population density.	200	Nicola Prevost	Yes	
4	12/12/2023	I added two additional visualisations and an OLS regression table.	180	Nicola Prevost	Yes	

4	12/13/2023	I developped two more Machine Learning models (SVM with different explanatory variables and K-Nearest Neighbours (KNN).	120	Nicola Prevost	Yes	
4	12/14/2023	I wrote the analysis for the Machine Learning models in the Juoyter Notebook and uploaded it to GitHub.	140	Nicola Prevost	Yes	
4	12/15/2023	I created a heatmap for pay parity across the world. I also tidied up the Machine Learning models and wrote the summaries for them.	120	Nicola Prevost	Yes	
4	12/16/2023	I did a complete run through of all the code and fixed parts that weren't working correctly.	180	Nicola Prevost	Yes	
		Total hours (excluding meetings):	55.42			