Programme	BSc (Hons) Data Scientist
Module name	Data Science Professional Practice
Schedule Term	Term 3
Student Reference Number (SRN)	BP0306578
Report/Assignment Title	Data Science Professional Practice Summative Assignment - Data Science Project & Impact Evaluation
Date of Submission *(Please attach the confirmation of any extension received)*	3rd September 2025

Declaration of Original Work:

- I hereby declare that I have read and understood BPP's regulations on plagiarism and that this is my original work, and that I have researched, undertaken, completed and submitted in accordance with the requirements of BPP School of Technology.
- I declare that where I have used any AI tools, it was for the following reasons (highlight as appropriate):
    - To create an assignment plan
    - To create a draft
    - To correct language errors
    - Other (please describe)
- I have copied & retained for BPP University's reference, all AI prompts used in the creation of AI content and all AI-generated responses in support of my assignment and attached relevant evidence in the appendices. I understand that I may be required to participate in a *viva voce,* where I will be questioned on any aspect of my assignment, including key concepts, theories, examples used, & any sources included.
- The word count, excluding contents table, bibliography and appendices, is **2978** words.

Student Reference Number: BP0306578 **Date**: 3rd September 2025

BPP School of Technology

Data Science Project

Table of Contents

How Preparedness for Civilian Life and Perceived Support Needs Relate to Loneliness and Income Among UK Veterans: A Cross-Tab Analysis of the 2022 Veterans' Survey

Data Science Project: How Preparedness for Civilian Life and Perceived Support Needs Relate to Loneliness and Income Among UK Veterans: A Cross-Tab Analysis of the 2022 Veterans' Survey

Executive Summary: This project analysed how UK veterans' preparedness for civilian life and perceived support needs relate to loneliness and income, using data from the 2022 UK Veterans' Survey (~28,000 responses). Preparedness was grouped into "Prepared", "Neither", and "Unprepared", and outcomes were explored through descriptive visualisations, chi-square tests, and ordinal logistic regression.

Findings showed preparedness is far more strongly linked to loneliness than income. Income distributions were similar across preparedness levels, with no significant associations. In contrast, unprepared veterans were substantially more likely to report loneliness: regression modelling estimated 67% higher odds for those "Neither prepared" and 3.5 times higher odds for those "Unprepared". Perceived support needs reinforced this, with lonely veterans most often identifying unmet mental health, counselling, and alcohol support.

Overall, the analysis highlights preparedness as a key determinant of social wellbeing, while financial outcomes appear shaped more by structural factors. Strengthening preparedness and psychosocial support is therefore critical to improving post-service transitions.

1. Introduction

This project investigated how UK veterans' self-reported preparedness and perceived support needs for civilian life relates to two outcomes: loneliness and income. It used publicly available data from the 2022 UK Veterans' Survey (Office for Veterans' Affairs and Office for National Statistics, 2023; ~28,000 veterans, Nov 2022–Feb 2023). By cross-tabulating loneliness and income against preparedness and perceived support needs, the analysis aimed to uncover how veterans' sense of preparedness and their identified support gaps relate to social and economic outcomes post-transition to civilian life.

2. Dataset Description

Respondents were aged 18+ and had served in the UK armed forces. This analysis used two pre-aggregated public datasets from the ONS website: one exploring links between loneliness and preparedness, and the other between income and preparedness. Data from England and Wales were weighted based on Census 2021 age profiles and presented with 95% confidence intervals (ONS, 2023).

3. Data Preparation

Each dataset was provided as an Excel file containing two disaggregated cross-tabulations: one for "Preparedness" and one for "Support Type". Both were transformed into long-form tables with a consistent structure. Terminology was standardised and "Prefer not to say" responses retained due to their notable proportions in some categories. Ultimately, a single merged table was avoided because preparedness and support type were disaggregated in the source data and not linked at row level, meaning there was no way to determine how individual support types mapped to specific preparedness categories.

4. Methodological Considerations

Preparedness was grouped as "Prepared", "Neither", and "Unprepared". Similarly, loneliness was simplified into three groups, and income into bands. All values were presented as weighted percentages and disclosure-protected by ONS. Variables used in both datasets were derived from standardised survey questions. Loneliness combined responses to multiple questions and income was self-reported gross income from all sources over 12 months. Ethical and governance concerns were mitigated by using anonymised public data.

5.1: Descriptive Visualisations:

The first stage of the analysis involved creating grouped bar charts and heat-maps to give a first look at how preparedness and perceived support needs relate to income and loneliness. This step provided a visual baseline before moving into more formal statistical testing involving chi-square and regression analysis.

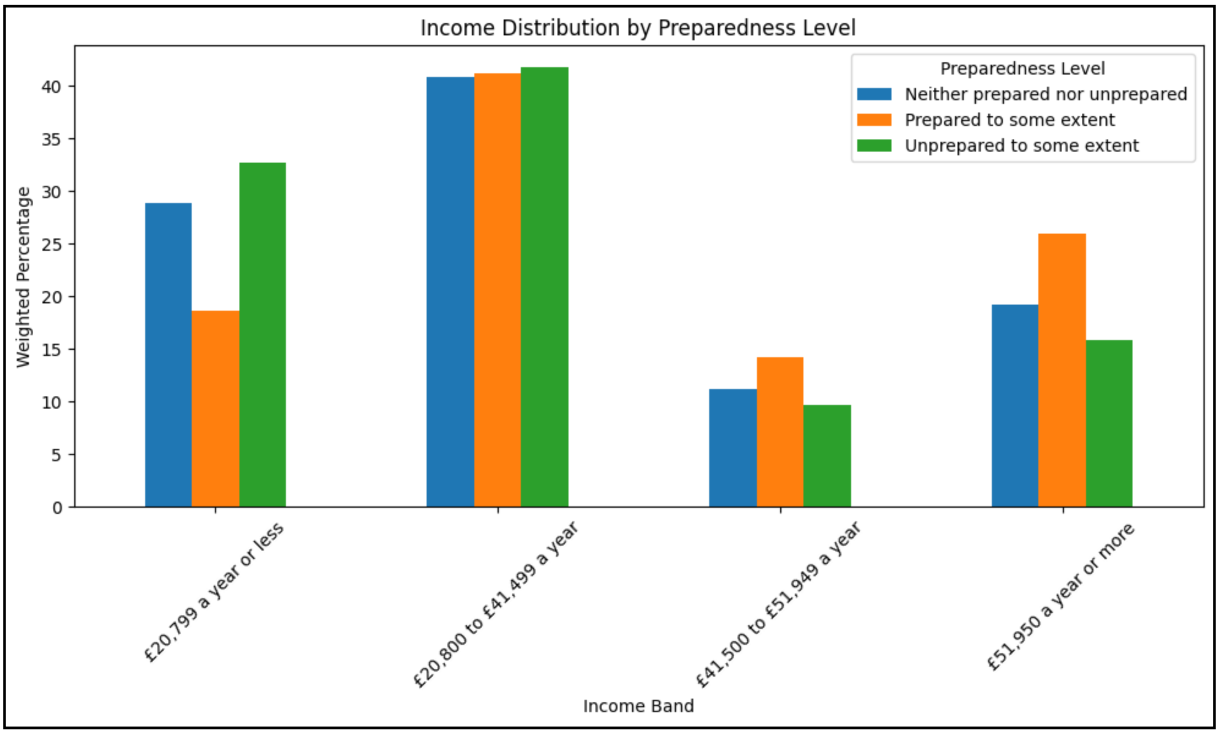


Figure 1. Income by Preparedness Level

Preparedness showed only weak differences in income distribution. Across all categories, the modal group was the £20,800–£41,499 band (~41%).

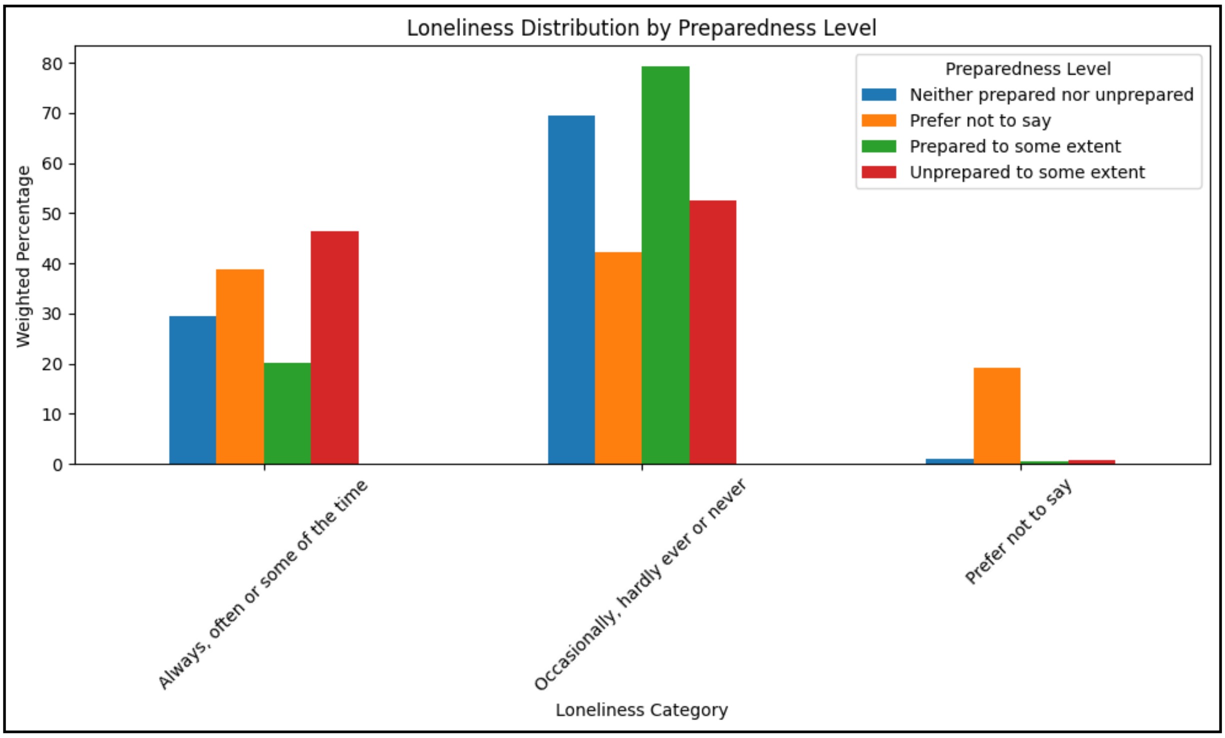


Figure 2. Loneliness by Preparedness Level

Among prepared veterans, around 80% reported low loneliness. This fell to ~53% among unprepared veterans, who were much more likely to report loneliness "always, often or some of the time."

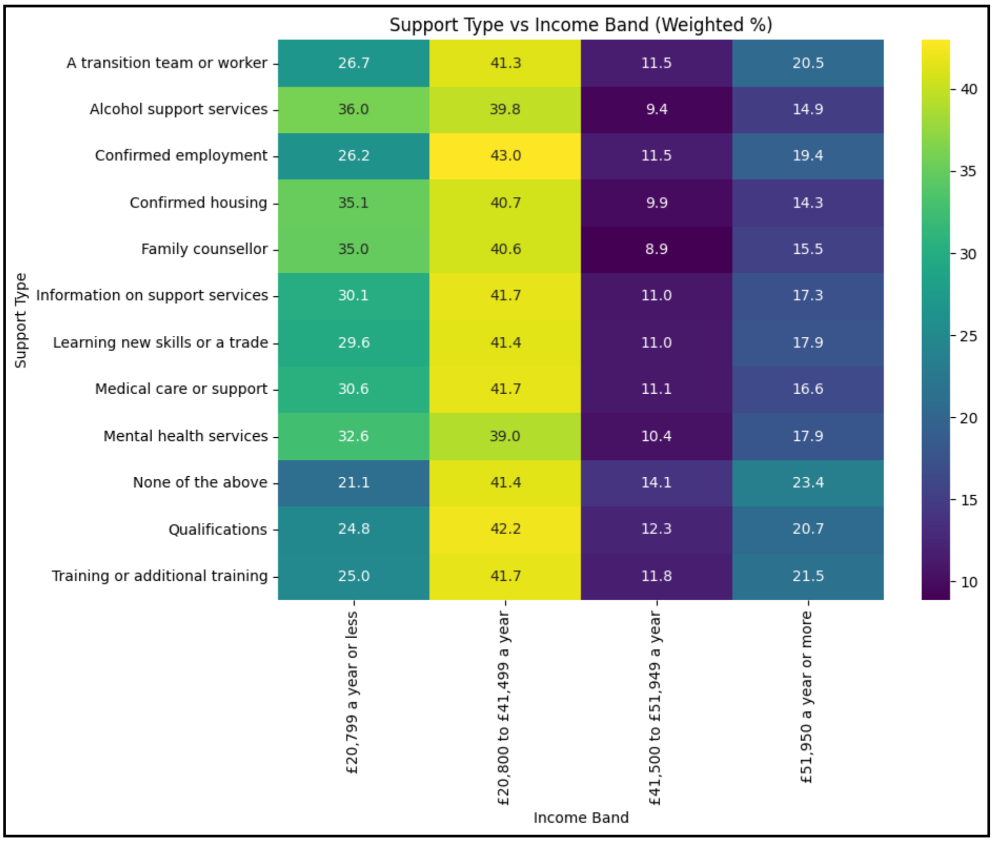


Figure 3. Support Type vs Income Band (Weighted %)

Lower-income groups more often cited housing, counselling, alcohol or mental health support; higher earners more often selected "None of the above".

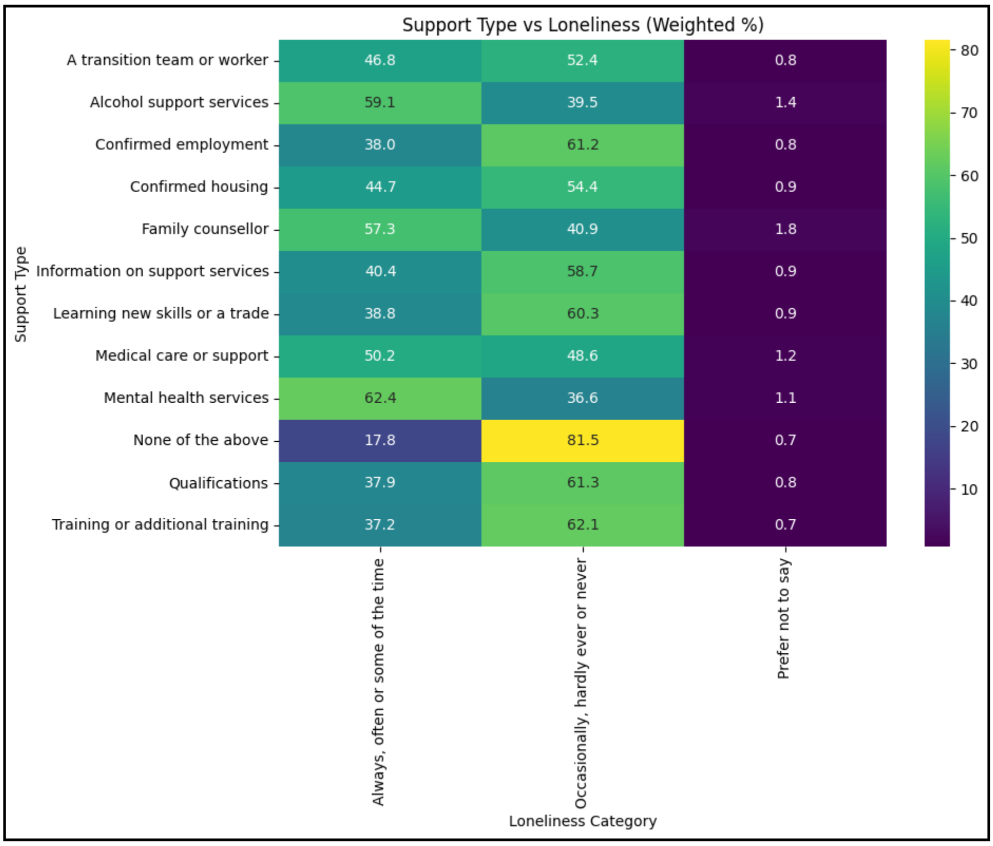


Figure 4. Support Type vs Loneliness (Weighted %)

Higher loneliness aligned with mental health support (62%), alcohol support (59%) and counselling (57%); low loneliness aligned with "None of the above" (82%). Sector reports identify elevated loneliness risks in parts of the Armed Forces community, consistent with these patterns (Royal British Legion, 2018).

These visuals show that preparedness is more closely tied to loneliness than income, while support needs cluster strongly around poorer social outcomes. They provide clear hypotheses for the statistical tests that follow.

5.2: Chi-Square:

The Data Analyst used chi-square tests of independence to examine whether preparedness and perceived support needs were associated with either income or loneliness. This test was well suited because it allowed the Data Analyst to see whether the distribution of one variable differs systematically across categories of another. Cramér's V was calculated to measure the strength of associations.

Using conventional benchmarks for Cramér's V (Cohen, 1988), income showed no association with either preparedness or support needs. By contrast, both preparedness and support needs were strongly linked to loneliness. Preparedness showed a moderate association, with unprepared veterans substantially more likely to report loneliness. Support needs were also significantly related to loneliness, though with a weaker effect size. Veterans reporting higher loneliness were more likely to highlight mental health services, counselling, alcohol support, and medical care as the types of help that would have improved their transition (see Appendix A, Table 1 for full chi-square results).

These results reinforce that preparedness and perceived support needs are more strongly linked to social wellbeing than to financial outcomes, which appear more influenced by structural drivers, such as skills, health, and regional labour markets (Forces in Mind Trust, 2021). The moderate effect sizes highlight that loneliness is not explained by preparedness alone, but the tests provide clear evidence of a real relationship. This motivated modelling loneliness only, with preparedness as the sole predictor, in the next stage of analysis.

5.3: Ordinal Logistic Regression

Ordinal Logistic Regression (OLR) was applied to the dataset within Google Colab to test whether preparedness could predict loneliness in a more nuanced way. OLR preserves the ranking of categories and is preferred when the dependent variable has ordered levels (Grace-Martin, 2016). OLR was thus suitable because it allowed the Data Analyst to estimate how much more likely veterans across the preparedness categories were to report loneliness. Models were estimated in Python using statsmodels' 'ordered logit implementation', recommended by Seabold and Perktold (2010).

The first attempt used the ONS tables directly. However, this produced results that were meaningless. This outcome reflected the limitations of the dataset insofar that the tables were aggregated, containing few rows, and leaving the model with insufficient information to estimate associations.

To address this, a pseudo-count expansion method was applied to create a dataset that reflected the proportions reported in the survey and contained enough rows to allow OLR to run on more data. Weighted percentages were scaled into counts (for example, 20% was treated as 200 out of 1000) and each row was expanded accordingly. Using pseudo-expanded data, the regression produced interpretable results.

Category	Odds Ratio (OR)	95% CI (Lower–Upper)
Neither prepared nor unprepared	1.67	1.36 – 2.05
Unprepared to some extent	3.49	2.86 – 4.25

Table 2. Odds ratios of frequent loneliness by preparedness for civilian life

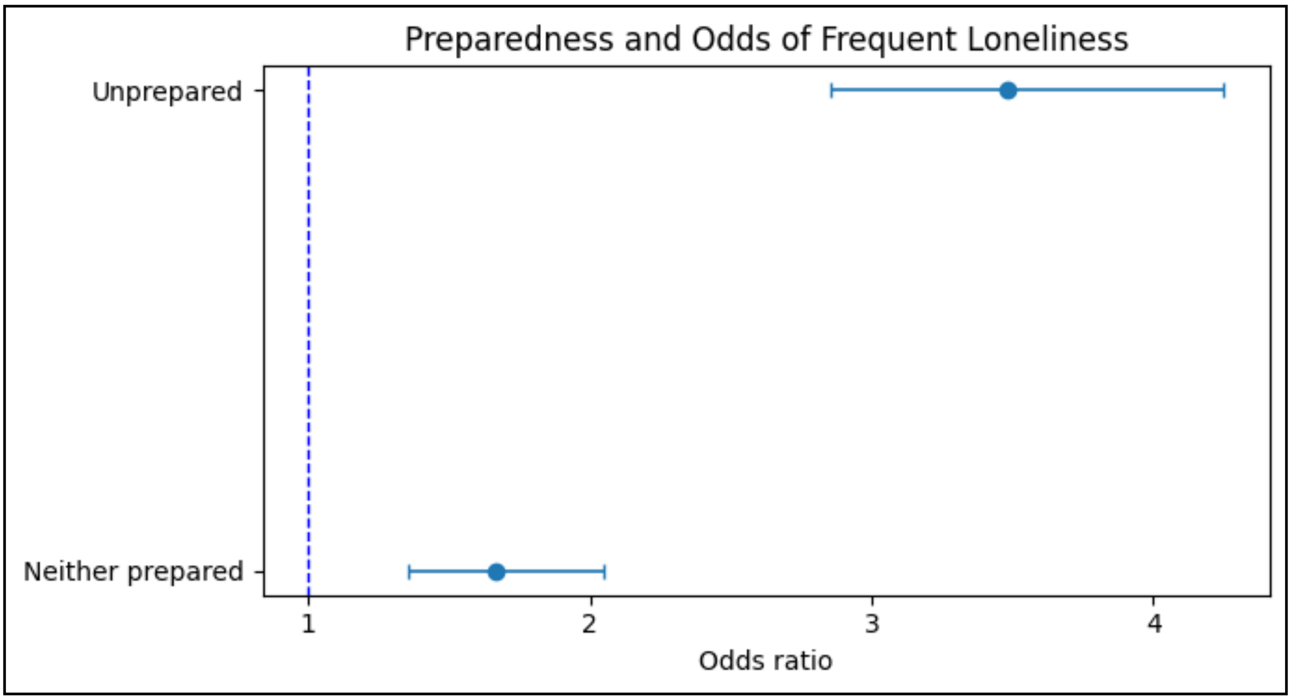


Figure 5. Forest plot of odds ratios (preparedness → loneliness)
Forest plot illustrating the relationship between preparedness for civilian life and frequent loneliness.

In the table, an odds ratio (OR) of **1.0** means a group is equally likely to report frequent loneliness as the baseline group (those who felt *prepared*). Values above 1.0 mean greater odds of loneliness. Veterans who were neither prepared nor unprepared had OR=1.67, meaning they were 67% more likely to report frequent loneliness compared with prepared veterans. Those who were unprepared had OR=3.49, meaning they were three and a half times more likely to report loneliness. Both results are statistically significant.

Figure X visualises these results. The vertical dashed line at 1 marks the point of "no difference". If a group's odds ratio falls exactly on this line, they are no more and no less likely to report frequent loneliness compared with the baseline group. The dots represent the estimated odds ratios for the "*Neither prepared*" and "*Unprepared*" categories, and the horizontal lines show the 95% confidence intervals. Both groups fall clearly to the right of 1, showing that they are more likely to report loneliness than prepared veterans. The "*Unprepared*" group sits further to the right, indicating the strongest likelihood of loneliness.

The chi-square test showed that preparedness and loneliness were connected. The regression quantified the size of that connection, showing that the impact of feeling unprepared is large. Whilst this stage of the analysis provides a precise estimate of how preparedness affects loneliness, limitations remain. The use of pseudo-counts is only an approximation and access to the raw survey responses would make it possible to build more robust models. Despite this, OLR confirms that preparedness for civilian life is strongly related to veterans' social wellbeing.

6. Conclusion

This project examined how preparedness for civilian life and perceived support needs relate to loneliness and income among UK veterans, using data from the 2022 Veterans' Survey.

The analysis showed that preparedness is more strongly linked to social outcomes than to financial ones. Income distributions were broadly similar across preparedness levels, with no significant associations in chi-square tests. In contrast, preparedness was significantly associated with loneliness, and regression modelling quantified the effect: veterans who were neither prepared nor unprepared were 67% more likely to report frequent loneliness, while those unprepared were around 3.5 times more likely. Perceived support needs mirrored this trend, with lonely veterans more likely to identify gaps in mental health, counselling, or alcohol support.

Overall, the project demonstrates that preparedness is a key factor in veterans' social wellbeing, particularly in reducing loneliness. The findings underline the importance of interventions that strengthen preparedness and provide psychosocial support in easing the transition to civilian life.

Bibliography A:

- Office for Veterans' Affairs and Office for National Statistics (2023) Veterans' Survey 2022: demographic overview and coverage analysis, UK. December. London: ONS. Available at https://www.ons.gov.uk/peoplepopulationandcommunity/armedforcescommunity/articles/veteranssurvey2022demographicoverviewandcoverageanalysisuk/december2023 (Accessed: [August 1st 2025])

- Royal British Legion (2018) Loneliness and social isolation support. London: Royal British Legion. Available at: https://www.britishlegion.org.uk/about-us/our-campaigns/campaigns/loneliness-and-social-isolation (Accessed 1st September 2025).

- Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Forces in Mind Trust (2021) Lifting Our Sights: Beyond 2030 — The Impact of Future Trends on the Transition of our Armed Forces Community. London: Forces in Mind Trust. Available at: https://www.fim-trust.org/project/beyond-2030-the-impact-of-future-trends-on-the-transition-of-our-armed-forces-community-from-military-to-civilian-life/ (Accessed 2nd September 2025).

- Grace-Martin, K. (2016) 'Logistic regression models for multinomial and ordinal variables', *The Analysis Factor*. Available at: https://www.theanalysisfactor.com/logistic-regression-models-for-multinomial-and-ordinal-variables/ (Accessed 2nd September 2025)

- Seabold, S. and Perktold, J. (2010) 'Statsmodels: Econometric and statistical modelling with Python', in van der Walt, S. and Millman, J. (eds.) Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, TX, 28 June–3 July. Available at: https://proceedings.scipy.org/articles/Majora-92bf1922-011 (Accessed 2nd September 2025).

Link to GitHub Repository: https://github.com/vastandinfinite95/BP0306578-DATA-SCIENCE-PROFESSIONAL-PRACTICE-SUMMATIVE-SUBMISSION

Appendix A: Supplementary Tables and Figures

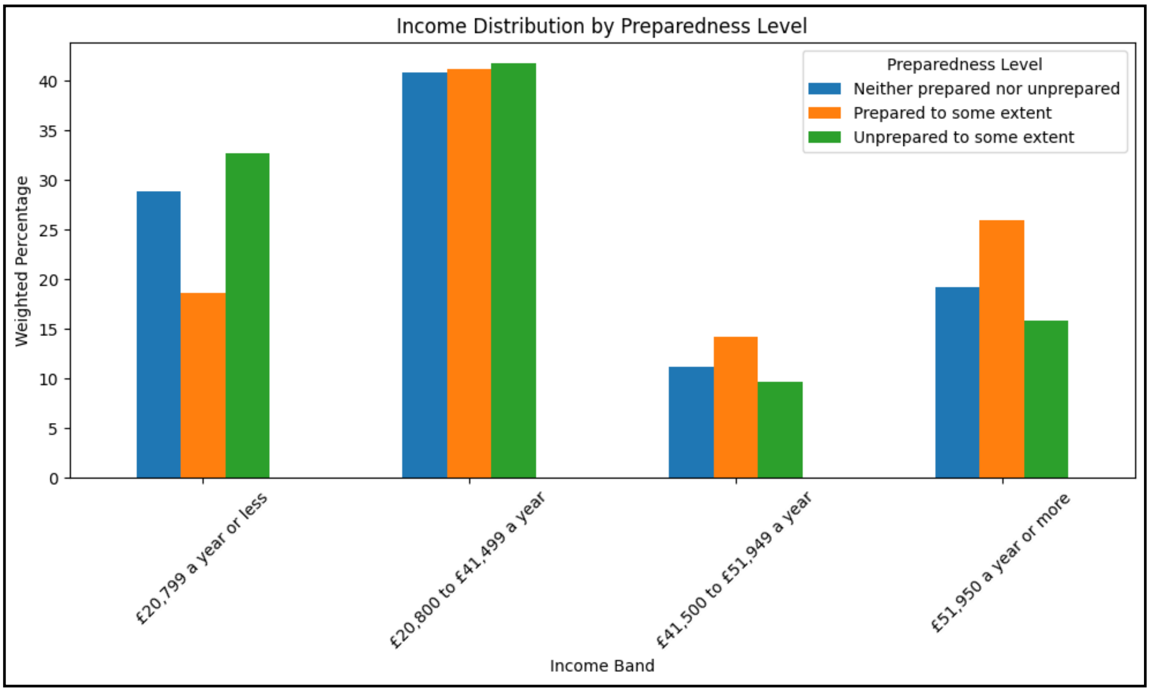


Figure 1. Income by Preparedness Level
Bar chart showing the distribution of veterans across income bands by preparedness category. Preparedness shows weak differences in income distribution.

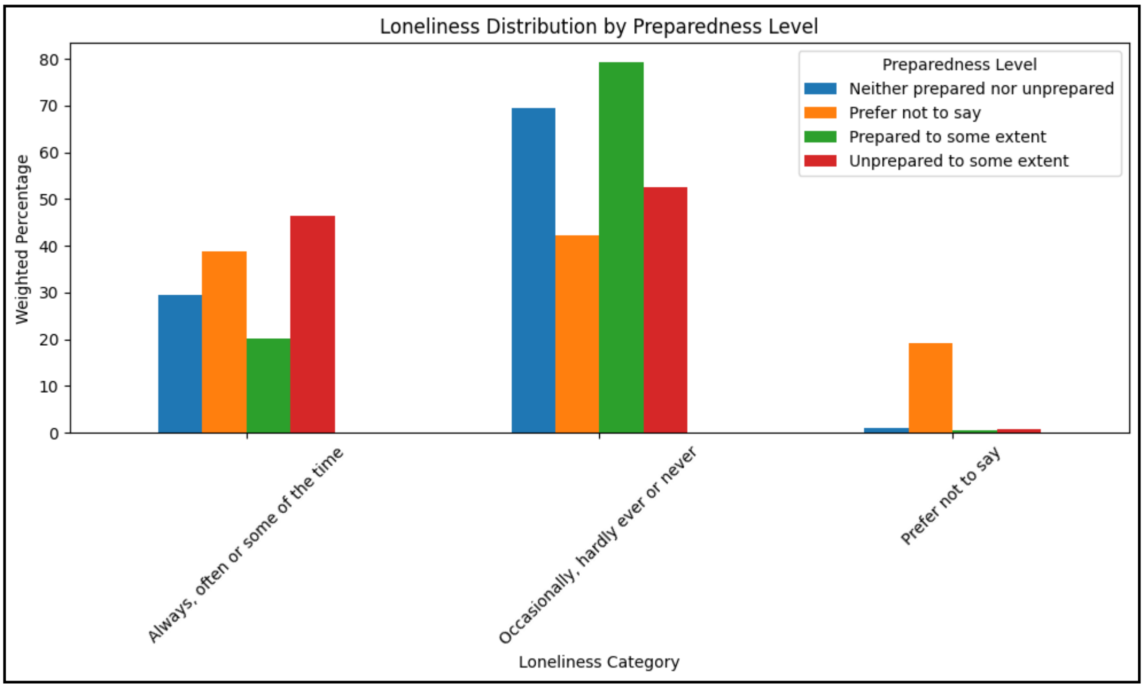


Figure 2. Loneliness by Preparedness Level
Bar chart showing loneliness outcomes across preparedness categories. Unprepared veterans are far more likely to report loneliness "always, often or some of the time."

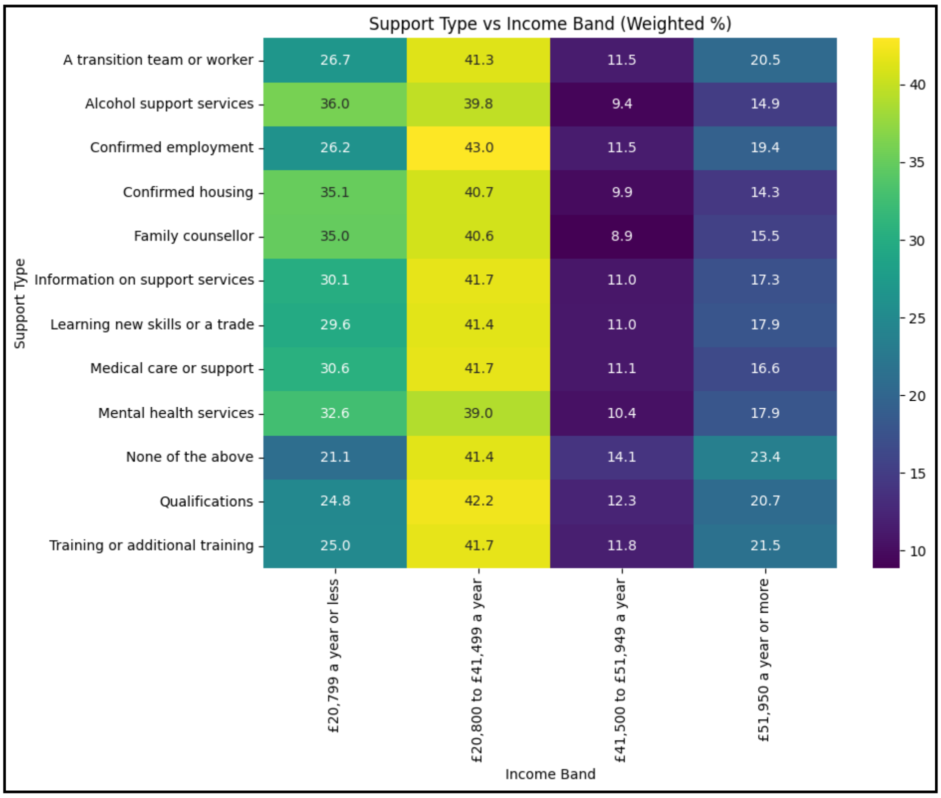


Figure 3. Support Type vs Income Band (Weighted %)
Heat-map illustrating income distribution across perceived support needs. Lower-income groups more frequently highlighted housing, counselling, alcohol or mental health support.

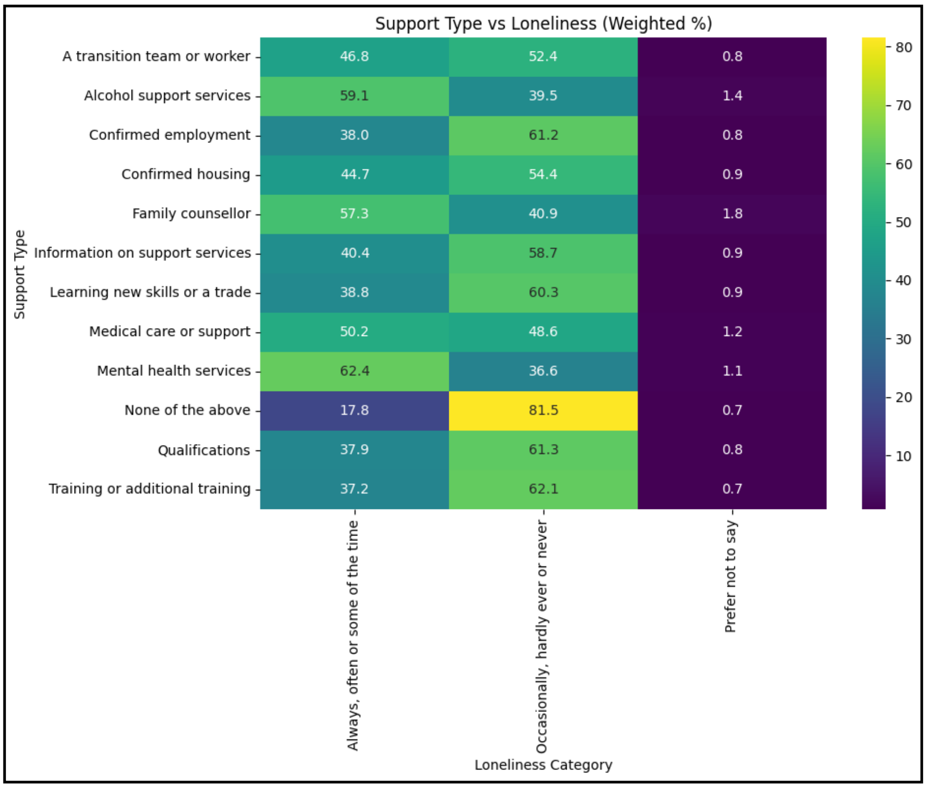


Figure 4. Support Type vs Loneliness (Weighted %)

Heat-map showing loneliness distribution across support needs. Higher loneliness aligns with mental health, alcohol, and counselling support, while low loneliness aligns with “None of the above.”

Test	χ^2 Statistic	df	p-value	Cramér’s V	Interpretation
Preparedness × Income Band	7.44	6	0.282	0.11	No statistically significant association. Income distribution is broadly similar across preparedness levels.
Preparedness × Loneliness	72.70	6	< 0.001	0.30	Significant association with a moderate effect size. Unprepared veterans are more likely to report frequent loneliness.
Support Type × Income Band	15.31	33	0.996	0.07	No statistically significant association. Perceived support needs are not meaningfully linked to income.
Support Type × Loneliness	69.22	22	< 0.001	0.17	Statistically significant but weak-to-borderline-moderate association. Veterans reporting loneliness are more likely to identify mental health/counselling-related supports.

Table 1: Summary of Chi-Square Test Results
This table provides the detailed statistical outputs referred to in Section 5.2.

Preparedness category	Odds Ratio (OR)	95% CI (Lower–Upper)
Neither prepared nor unprepared	1.67	1.36 – 2.05
Unprepared to some extent	3.49	2.86 – 4.25

Table 2. Odds ratios of frequent loneliness by preparedness for civilian life (from OLR)
Results of the ordinal logistic regression referred to in Section 5.3.

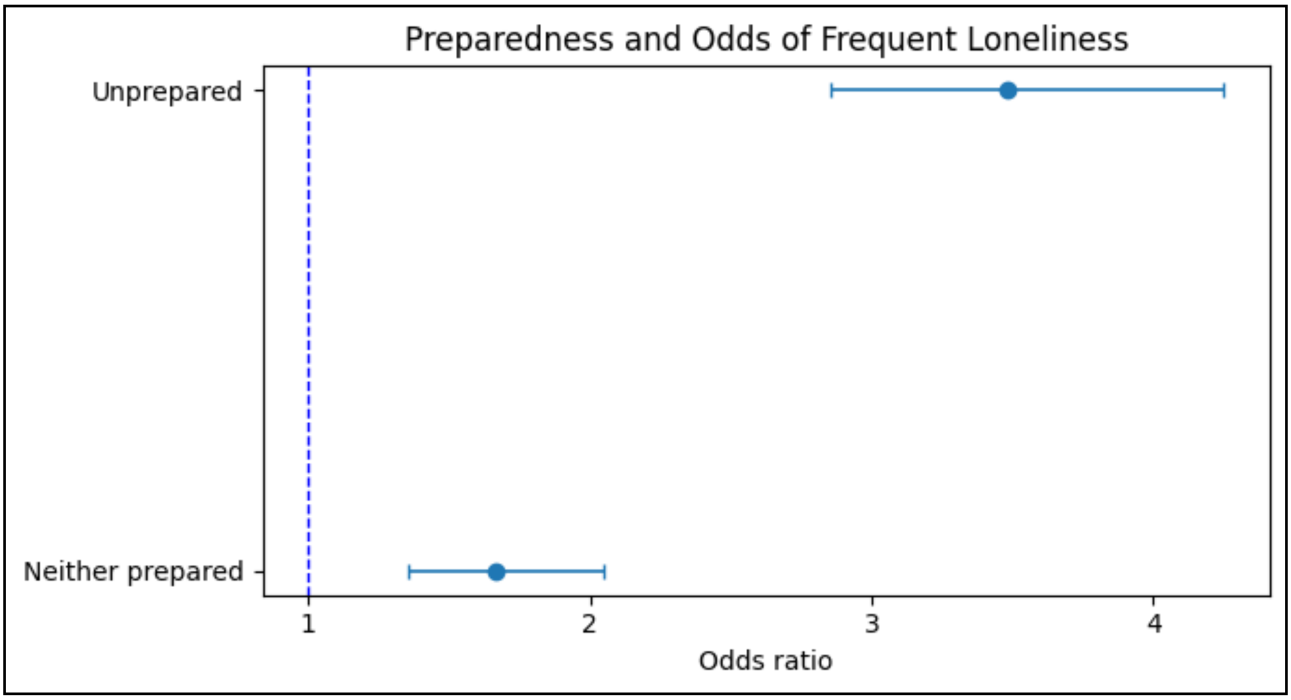


Figure 5. Forest plot of preparedness and odds of loneliness
Forest plot illustrating the regression results from Table 2. Both “Neither prepared” and “Unprepared” groups lie clearly to the right of the no-difference line (OR=1), with the “Unprepared” group showing the strongest effect.

Impact Evaluation

Table of Contents

Impact Evaluation: What Works When: Timing One-to-One Attendance at LCF with Logistic Regression Heat-maps

Executive Summary: Graduate Futures (GF) needed evidence on when to schedule one-to-ones so bookings convert to attendance. The Data Analyst used routine Moodle exports within GF's governed data flow (SharePoint and Excel) and ran logistic regression in Google Colab, producing Hour x Day and Student Year x Day heat-maps.

Findings showed a clear late-morning to early-afternoon, mid-week band of stronger attendance, with weaker late-day windows. First Years were more sensitive to day choice and later years were more tolerant. The visuals were embedded in the Academic Year Dashboard so consultants and operations could align rotas and schedules. Governance followed data minimisation and restricted storage, with "N/A (unknown)" handled transparently.

Model performance was moderate but sufficient to rank time windows. The repeatable pipeline will refresh each term, growing the dataset and stabilising estimates.

Recommendations include publishing a termly Best-Slot Matrix from the latest model, tracking before/after no-show rates, and extending the model with additional features such as lead time, reminders sent, and session length.

1. Project Background

Graduate Futures (GF) at London College of Fashion delivers a wide range of student support. The core provision includes tailored one-to-one consultations across Careers (career planning, CV's, applications) and Enterprise (freelancing, start-up advice). These consultations are complemented by a broad portfolio of webinars, workshops, and in-person events.

Earlier analysis established what features of the core provision tend to convert student bookings to student attendance. For example, one-to-ones outperformed webinars. What the GF team lacked was evidence about when to schedule consultations to maximise student follow-through. This project addressed that gap; the Data Analyst analysed timing effects and translated them into guidance that could be applied directly to scheduling strategies of GF consultants.

The work sat inside GF's usual Moodle → SharePoint → Excel flow (See Figure 1) and added a Colab step to produce logistic-regression heat-maps (See Figure 2). GF makes hundreds of scheduling decisions each term, and every unattended one-to-one is a lost 30 to 60 minute slot. These outputs, on display in the Academic Year Dashboard (AYD), derive high impact potential, as even modest lifts in attendance for one-to-ones translate into less wastage of GF resources and more students receiving support.

This project operationalised a "future improvement" proposed in earlier work and shifted to a temporal view of what works when.

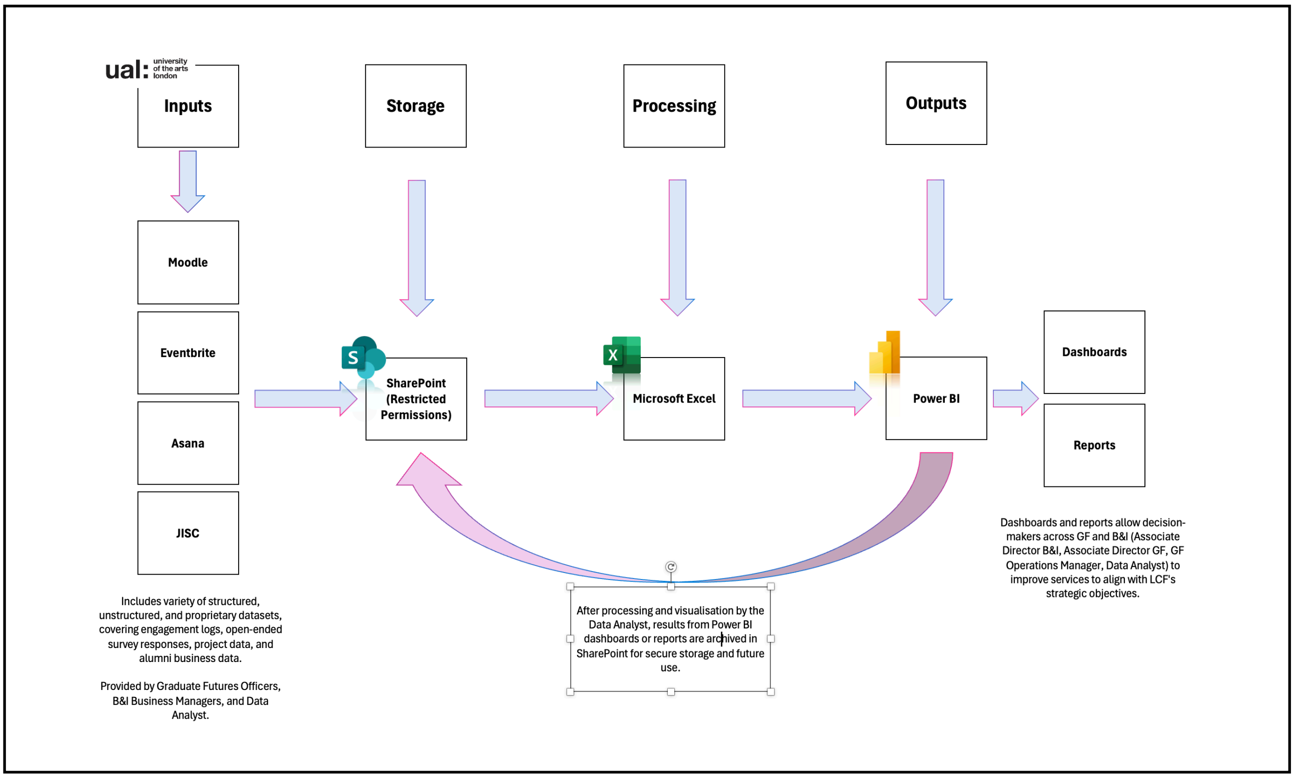


Figure 1. Graduate Futures data flow (inputs → SharePoint → Excel/BI → outputs). Standard governed flow used across GF. This project adds a modelling branch in Google Colab.

2. Approach & Analysis

The workflow began with routine Moodle exports of one-to-one data. In Excel the exports were combined, identifiers removed, times parsed, Hour and Day derived, types fixed, and duplicates removed. The full path from source to analysis is summarised in Figure 2, which extended the established GF data flow mentioned in the previous section by adding a modelling branch in Google Colab.

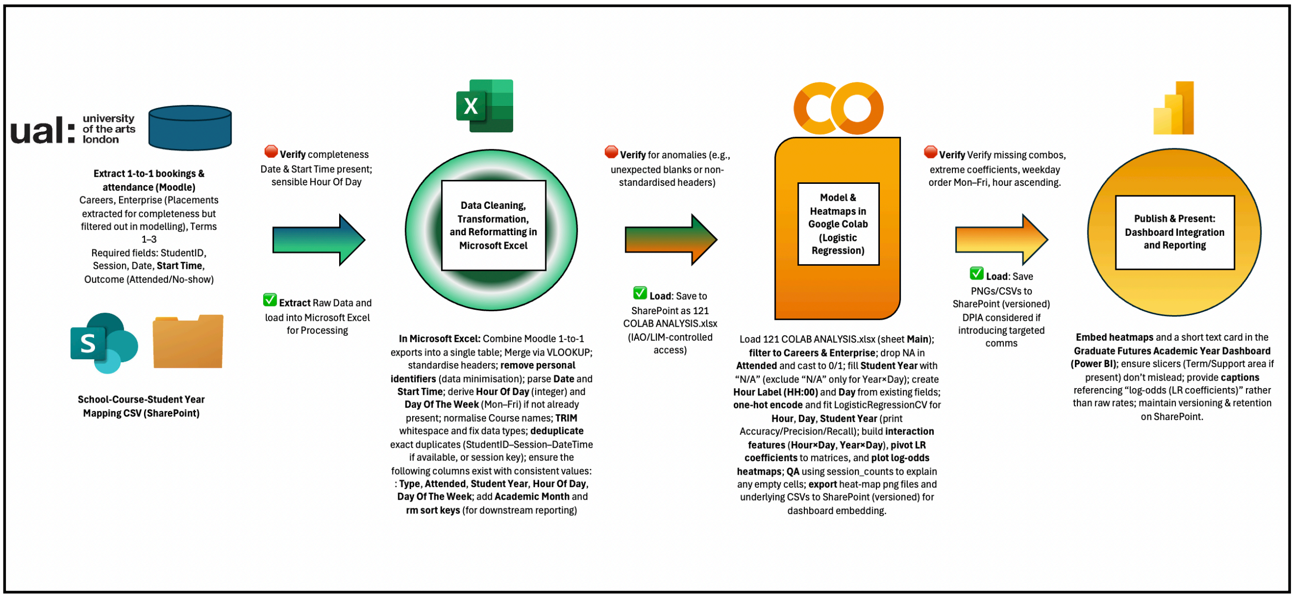


Figure 2. ETL + Modelling workflow for 1-to-1 attendance timing. Excel ETL produced an analytics ready sheet; Colab fitted logistic regression and returned heat-maps; outputs were embedded in the Graduate Futures Academic Year Dashboard.

In Colab, the Data Analyst read the Excel sheet into pandas and filtered to Careers and Enterprise consultations. Using a notebook environment (Jupyter/Colab) combined code, narrative and results in one auditable document, improving reproducibility and sharing across the team (Rule et al., 2019). Rows without an attendance value were excluded, and 'Attended' converted to 1 for attended, 0 for no-show. Missing Student Year was labelled 'N/A (unknown)'; it appeared in summaries and was handled separately in the Year x Day visual.

Before looking at combinations of timing, the Data Analyst ran simple checks one factor at a time (hour, day, year), revealing how each factor effected the log-odds of attendance. This confirmed that each factor carried insight individually.

Logistic regression was appropriate because the outcome was binary, the coefficients are interpretable for non-technical audiences, and interactions (for example, Hour x Day and Student Year x Day) could be modelled to test combined timing effects (Hosmer, Lemeshow and Sturdivant, 2013). The modelling balanced the classes so that attendance and no-shows were weighed fairly. This reduced the risk of change patterns and helped the model pay attention to the less common outcome. A fixed random setting was used so the analysis was repeatable.

Scheduling decisions, however, depended on combinations. For that reason, the analysis examined two interactions: Hour x Day (for daily windows that consistently converted) and Student Year x Day (for cohort specific timing). Each combination was compared with the model's baseline and then pivoted into a matrix so the results could be read as a heat-map. The colour scale used a neutral midpoint so warm cells indicated a positive effect and cool cells a negative effect. Coefficients were small in magnitude (roughly -0.23 to +0.21 on the log-odds scale), which is typical for behavioural attendance data.

The two visuals were presented as Figure 3 (Hour x Day) and Figure 4 (Student Year x Day).

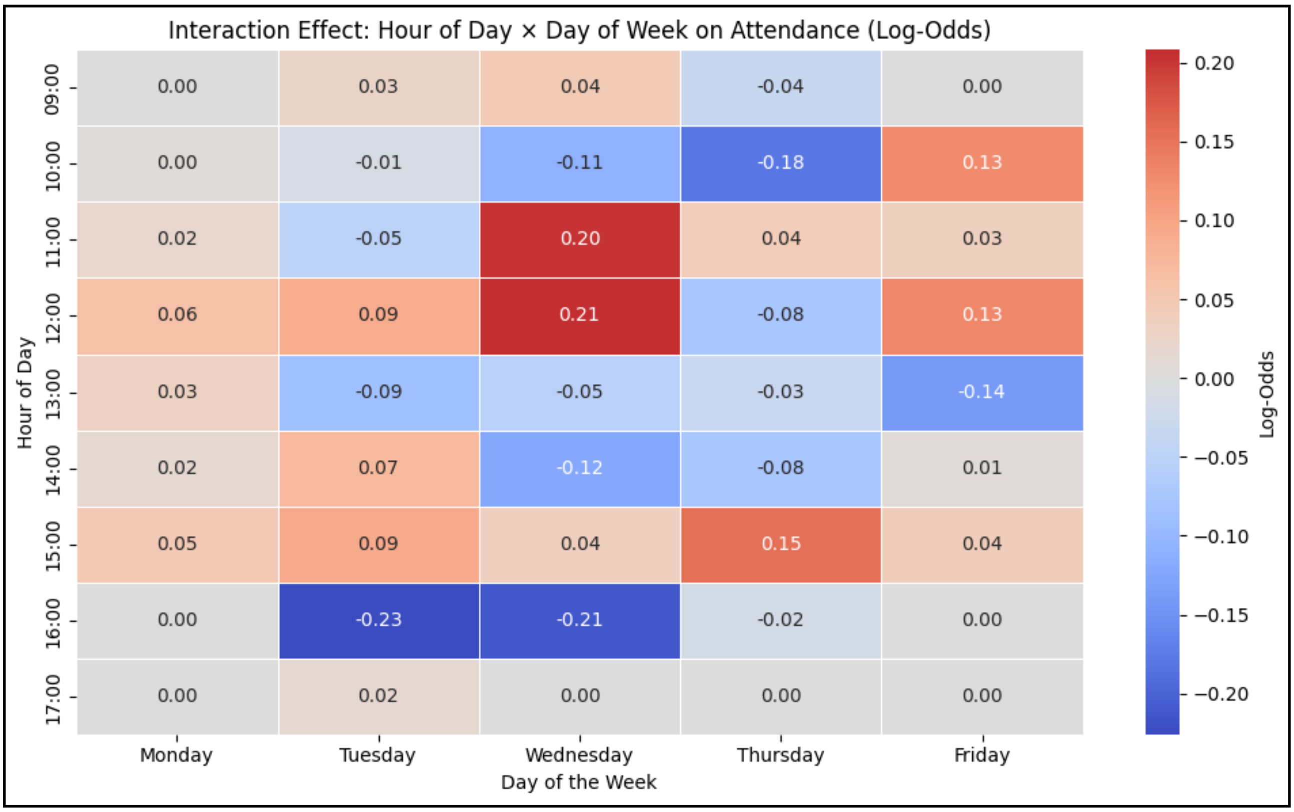


Figure 3. Interaction heat-map: Hour of Day x Day of Week (model effect, neutral midpoint). Blank cells indicate combinations with no underlying sessions.

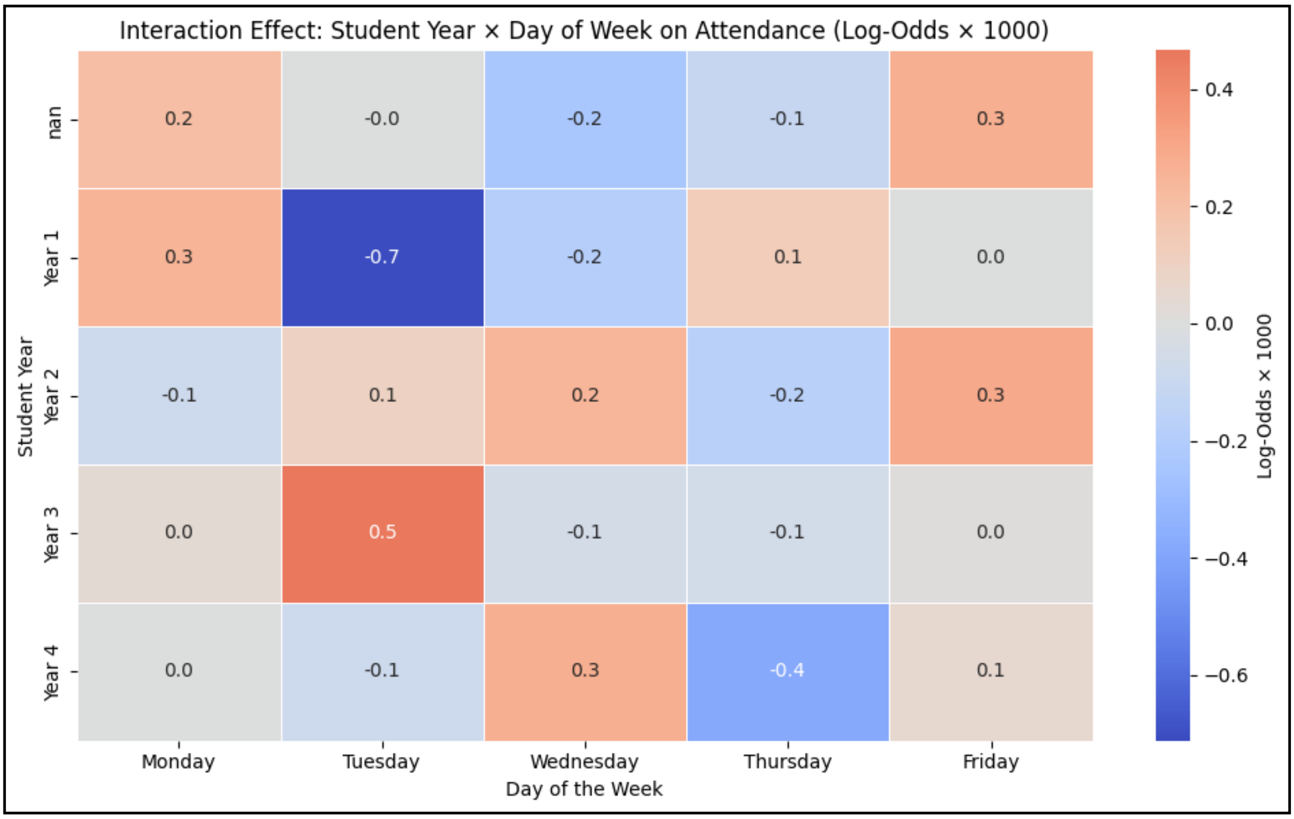


Figure 4. Interaction heat-map: Student Year x Day of Week (model effect, neutral midpoint; values scaled ×1000 for readability). The "N/A (unknown)" cohort was displayed as a separate row.

Model performance was checked and summarised. Hour x Day achieved accuracy 0.502, precision 0.687, recall 0549; Year x Day achieved accuracy 0.512, precision 0.725, recall 0.527. The interaction models showed moderate discrimination, sufficient for the operational goals: ranking time windows, grounding scheduling choices in evidence, and indicating when one-to-ones were most likely to be attended, rather than predicting individual outcomes.

Governance followed GF practice throughout with a focus on data minimisation and pseudo-IDs, SharePoint storage under IAO/LIM oversight, and DPIA consideration if targeted communications were introduced. The intention was to improve access by offering earlier slots or different days to cohorts with weaker follow-through, rather than to deprioritise them. Finally, the heat-maps and the supporting tables were saved back to SharePoint and embedded in the AYD. Captions made clear the colours showed the model's relative effect rather than raw attendance rates, so the visuals could be used responsibly. Clear labelling and explanation align with the UK Government Data Ethics Framework principles of transparency and fairness, helping teams use analytics responsibly in operational decisions (CDDO, 2020).

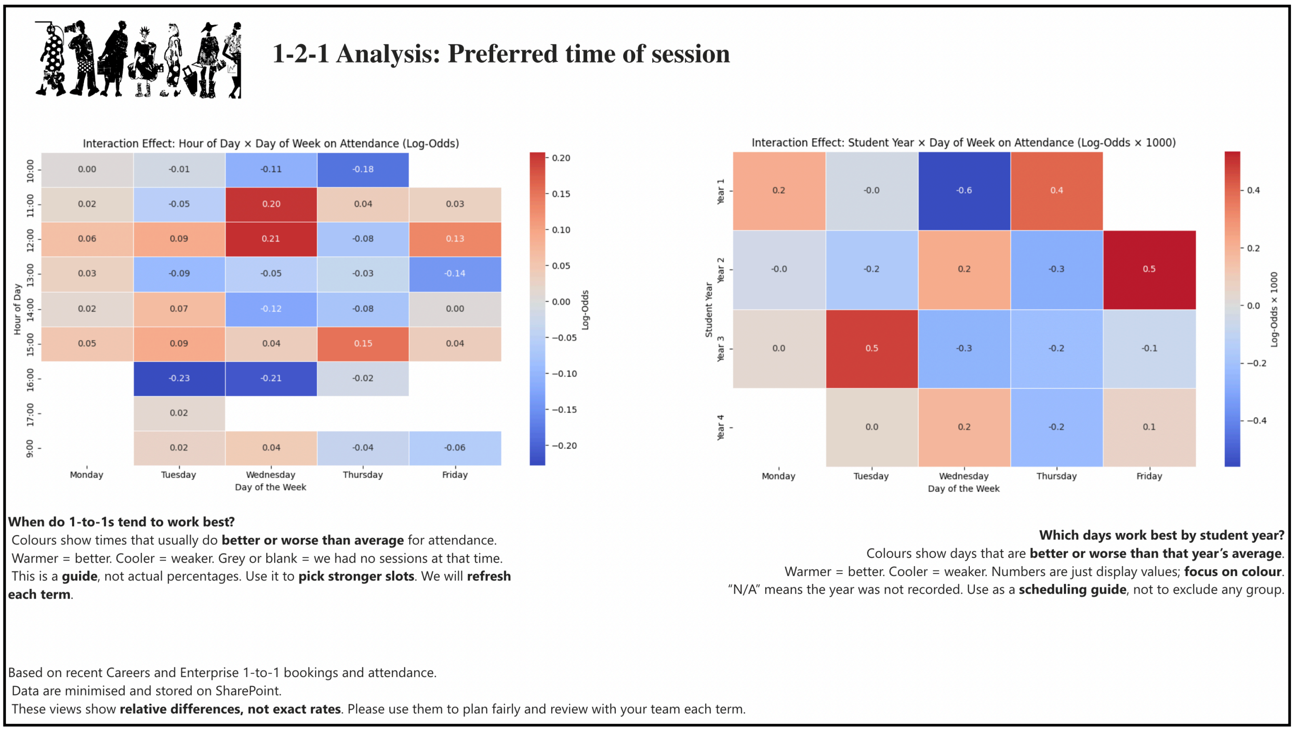


Figure 5. Dashboard integration. Heat-maps were surfaced within the Academic Year Dashboard with captions explaining that values show the model's relative effect, not raw rates.

3. Results & Impact

The analysis translated directly into scheduling decisions. Figure 3 showed late-morning to early-afternoon preference mid-week. Figure 4 showed that First Years were especially sensitive to day choice, whereas students in later years were more tolerant. The Data Analyst worked with Operations and the Head of Graduate Futures Consultants to move one-to-ones for First Years away from weaker windows, for example, late Tuesday's and Wednesday's, and cluster general one-to-ones into the strongest bands, late morning, mid-week.

Colleagues reported an immediate shift in planning conversations. As the Head of Graduate Futures Consultants noted:

"The heat-maps changed rota meetings from 'when do we think this might work?' to 'here's where students are most likely to show up'. We moved Year 1 sessions into the stronger day bands and concentrated generic slots late morning. The days felt slightly calmer, consultants had fewer gaps, and students were seen sooner."

This shift toward using embedded analytics in routine planning reflects wider evidence that organisations which operationalise analytics make faster, better-aligned decisions (Davenport and Harris, 2007).

Embedding the visuals in the **AYD** (Figure 5) gave GF a shared evidence base. Consultants used the heat-maps to pick times for sessions and Operations used them to design the delivery rotas for the next academic year. The analysis supported equity of access. By making cohort sensitivity visible, GF could offer earlier slots or different days to groups with weaker-follow through without reducing overall availability. Finally, the method created a repeatable pattern to be used throughout the next academic year and beyond.

Stage	Tool
1. Export	Moodle
2. Tidy-up	Excel
3. Model & Heat-maps	Colab
4. Publish	Power BI Dashboard

Table 1. Simple and repeatable analysis pipeline

Looking ahead, the repeatable pipeline meant insights could be refreshed each term and rolled into annual planning, making the analysis part of GF's established data flow. While the interaction models in this run showed moderate discrimination, as routine exports accumulate each term, the training dataset will grow, and sparse timetable edges will be filled, meaning coefficients should stabilise and improve. As Halevy, Norvig and Pereira observe, "simple models and a lot of data trump more elaborate models based on less data." (Halevy, Norvig and Pereira, 2009).

4. Recommendations

To convert the analysis into lasting operational value, the Data Analyst prepared the following proposals for discussion with GF.:

1. Publish a termly Best-Slow Matrix (from that term's model)

Publish a one-page Hour x Day "best-slot matrix" at the start of each term, generated from that term's logistic regression interaction results, and embed it in scheduling or rota templates. The page should:

- Rank Hour x Day windows overall and by cohort where useful.
- Set a default by prioritising top-ranked windows for scheduling and avoid the lowest-ranked unless capacity requires.
- Embed the matrix in rota packs and the AYD. Its contents will change each cycle as fresh data is analysed.

2. Measure outcomes and iterate (before/after)

Add dashboard tiles for no-show rate by Hour x Day, before versus after timetable changes, plus attended hours recovered (sum of 30 to 60 minute slots saved). Review monthly and at term's end.

3. Extend with further interaction features

Add new interaction features that the team can influence, such as lead time from booking, reminder sent or seen, and session length.

Finally, in terms of governance also, the Data Analyst will recommend to move routine transformations to Power Query or a small python script, for example, to validate weekday/hour fields, and continue to keep assets on SharePoint with IAO/LIM permissions.

Bibliography B:

- Rule, A., Birmingham, A., Zuniga, C., Altintas, I., Huang, S.-C., Knight, R. et al. (2019) 'Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks', PLOS Computational Biology, 15(7). Available at: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007007 (Accessed: 2nd September 2025).

- Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression. 3rd edn. Hoboken, NJ: Wiley. Available at: Wiley Online LibraryWiley (Accessed: [1st September 2025]).

- CDDO (2020) *Data Ethics Framework*. London: Central Digital and Data Office. Available at: https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020 (Accessed: 2nd September 2025).

- Davenport, T.H. (2006) 'Competing on Analytics', Harvard Business Review, January. Available at: https://hbr.org/2006/01/competing-on-analytics (Accessed: [28th August 2025]).

- Halevy, A., Norvig, P. and Pereira, F. (2009) 'The unreasonable effectiveness of data', IEEE Intelligent Systems, 24(2), pp. 8–12. Available at: https://research.google.com/pubs/archive/35179.pdf (Accessed: [28th August 2025]).

Link to GitHub Repository: https://github.com/vastandinfinite95/BP0306578-DATA-SCIENCE-PROFESSIONAL-PRACTICE-SUMMATIVE-SUBMISSION

Appendix B: Supplementary Tables and Figures

Figure 1. Graduate Futures data flow (inputs → SharePoint → Excel/BI → outputs). Standard governed flow used across GF. This project adds a modelling branch in Google Colab (see Figure 2).

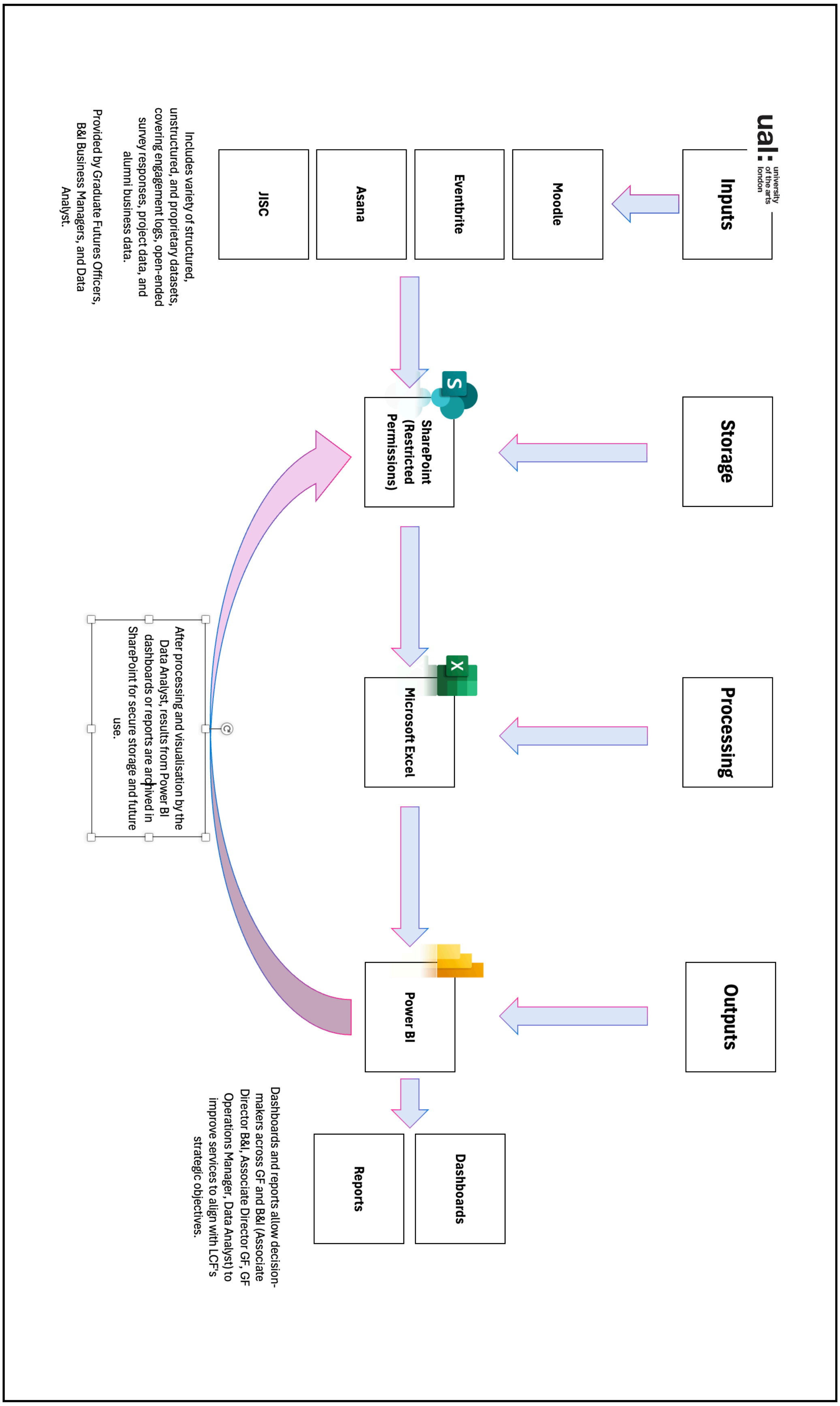
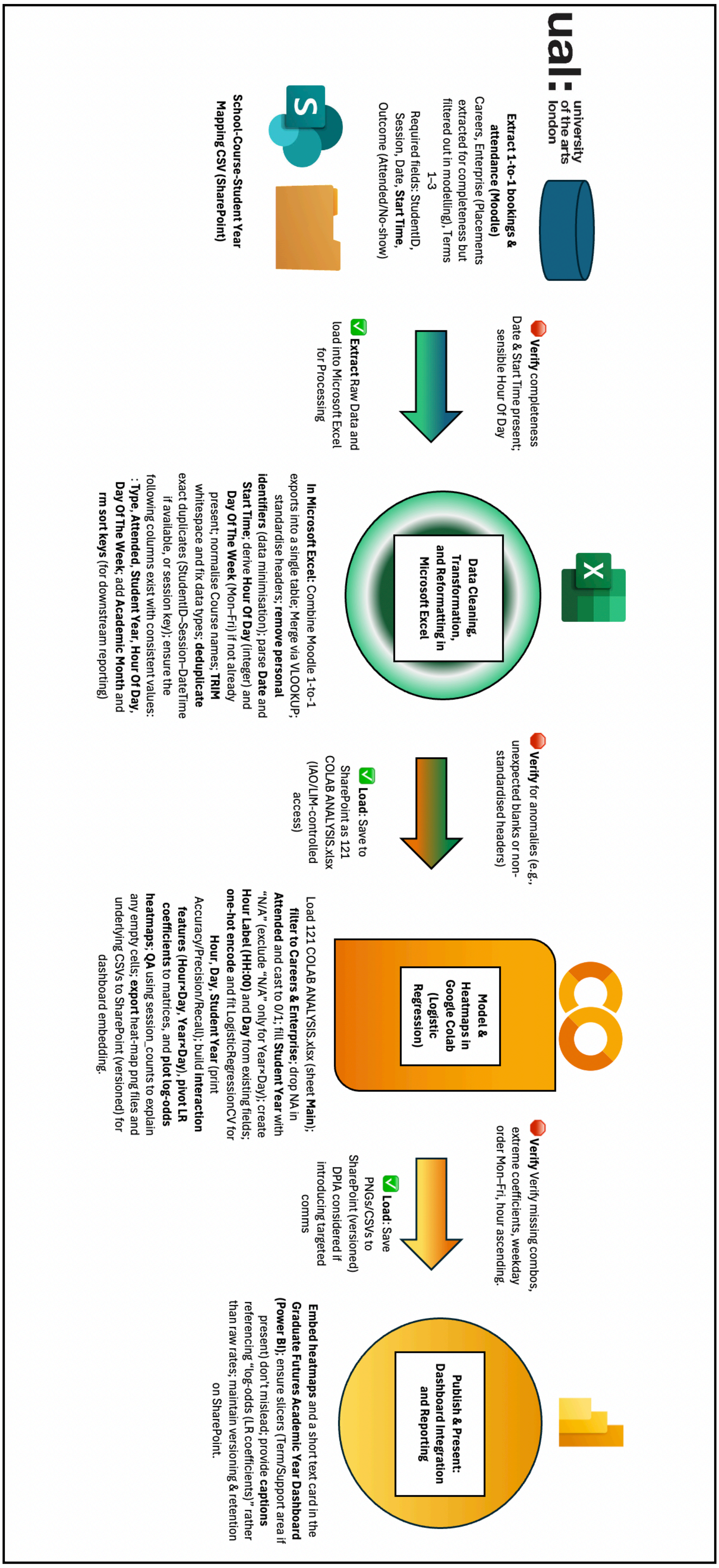


Figure 2. ETL + Modelling workflow for 1-to-1 attendance timing. Excel ETL produced an analytics ready sheet; Colab fitted logistic regression and returned heat-maps; outputs were embedded in the Graduate Futures Academic Year Dashboard.



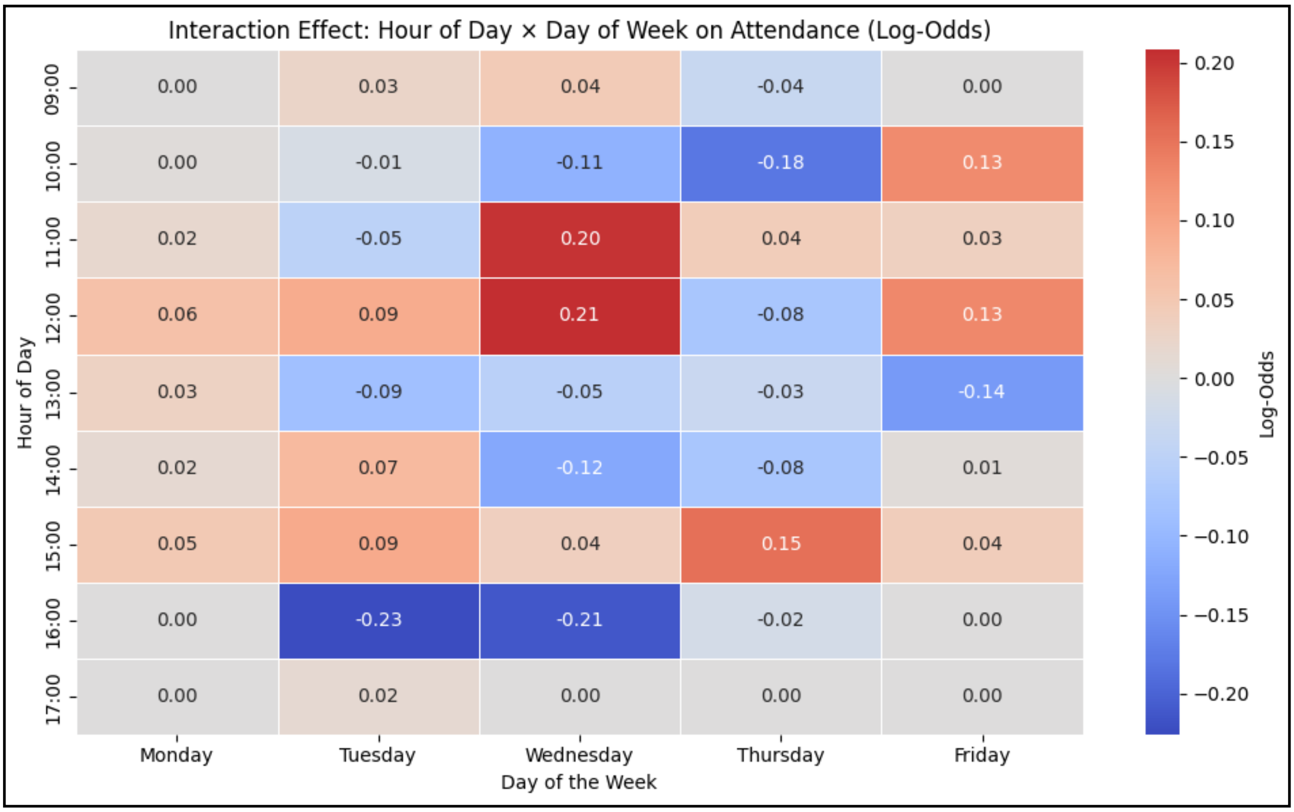


Figure 3. Interaction heat-map: Hour of Day × Day of Week (model effect, neutral midpoint). Blank cells indicate combinations with no underlying sessions.

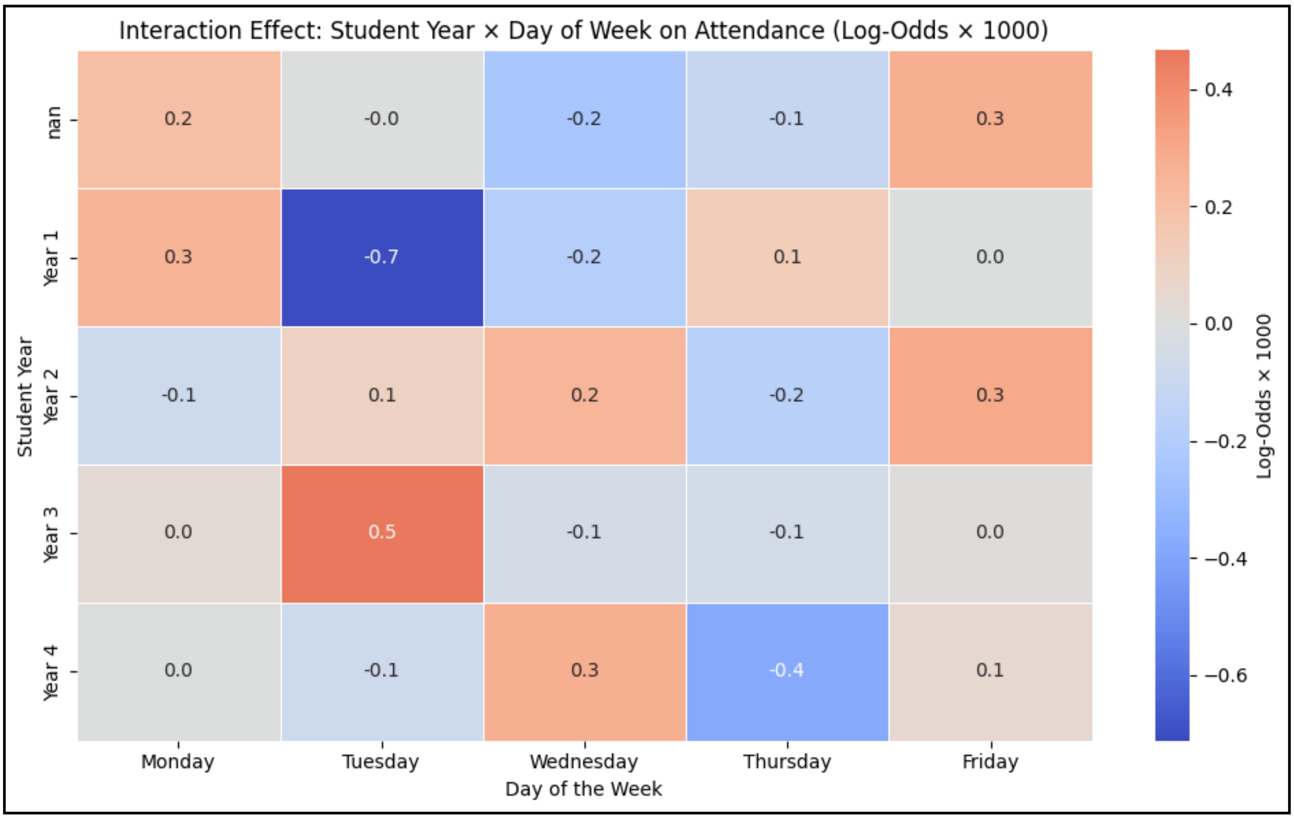
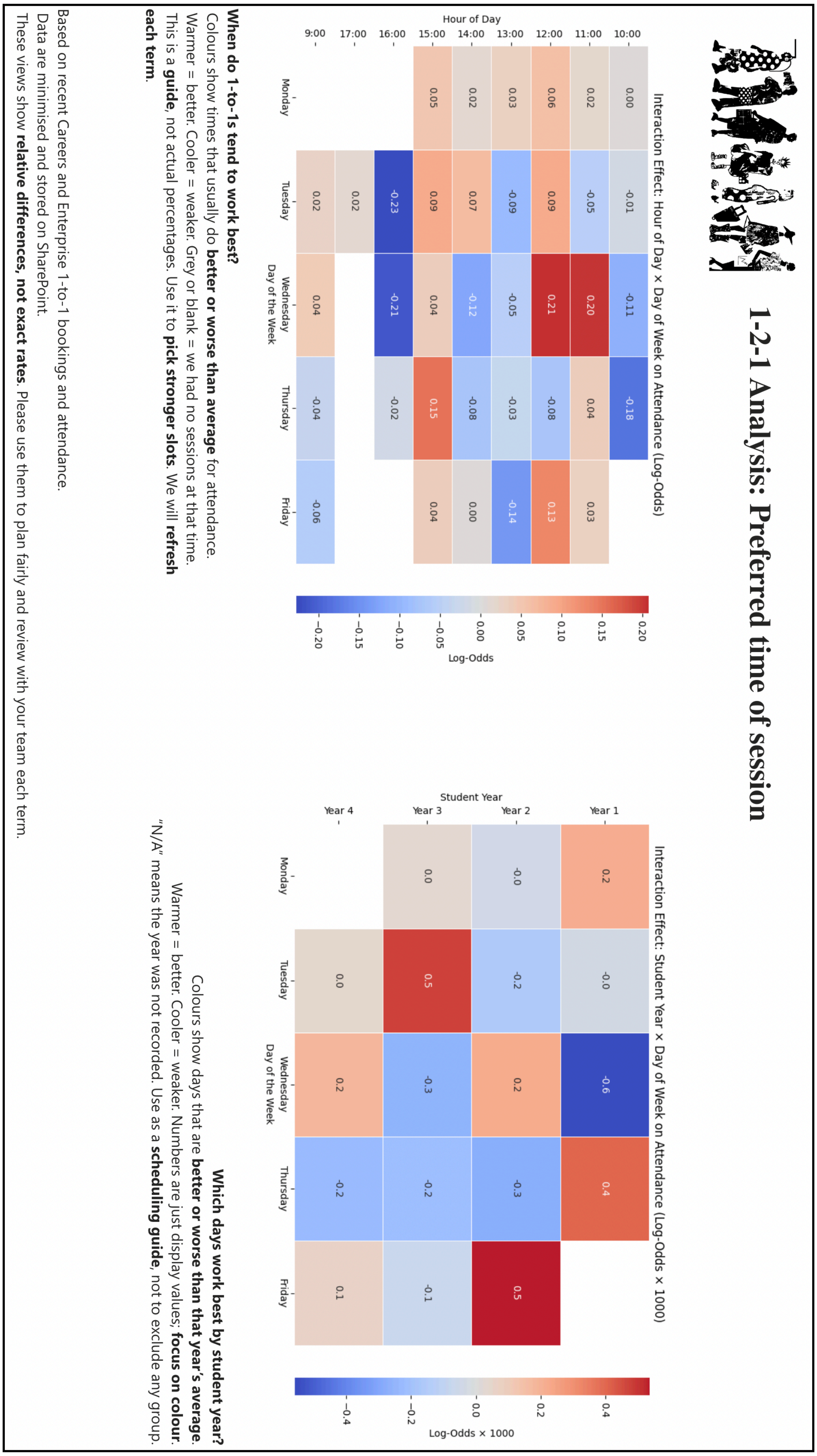


Figure 4. Interaction heat-map: Student Year × Day of Week (model effect, neutral midpoint; values scaled ×1000 for readability).

Figure 5. Dashboard integration. Heat-maps were surfaced within the Academic Year Dashboard with captions explaining that values show the model's relative effect, not raw rates.



Stage	Tool
1. Export	Moodle
2. Tidy-up	Excel
3. Model & Heat-maps	Colab
4. Publish	Power BI Dashboard

Table 1. Simple and repeatable analysis pipeline