|  |  |
| --- | --- |
| **FRONT SHEET**  **Individual Coursework** | |
| **CANDIDATE NUMBER (C-NUMBER)** | 0400248 |
| **MODULE NAME** | Data and Decision Making |
| **WORD COUNT** | 2440 |
| **SUBMISSION DATE** | 09/22/2025 |
| **DECLARATION**  I certify that this assessment submission is entirely my work and I have fully referenced and correctly cited the work of others, where required. I also confirm the contents of my submission have not been generated by a third party, or through an Artificial Intelligence generative system\*.  I have read the Student Discipline Regulations ([Student Discipline Regulations](https://www.law.ac.uk/globalassets/13.-media--doc-repo/08.-policies/pdf_policies_student-discipline-regulations.pdf)) and understand any Assessment Related Offence/ Academic Misconduct may result penalties being applied.  By submitting this assessment submission, I am confirming that I am fit to sit according to the Assessment Regulations.  I declare that:   1. This is my own unaided work.   Yes  No   1. The word count stated by me is correct.   Yes  No   1. I’m happy for my work to be retained on the Elite repository and made available to staff and future students\*\*   Yes  No  \*Please note that all the assignments are submitted to Turnitin.  \*\*Please note personal information (such as names) will be deleted. | |
| |  | | --- | | **Instructions to candidates:**   1. **Please complete this cover sheet by entering your Candidate Number, Module Name, Word Count, and Submission Date.** 2. **You must NOT use your NAME on this cover sheet or on any part of your coursework.** | | |

Data-Driven Decision Making Report

Table of Contents

[Executive summary 5](#_Toc209439235)

[1. Introduction 6](#_Toc209439236)

[2. Main analysis and tasks 6](#_Toc209439237)

[2.1 Topic, business decision, and analytical framing 6](#_Toc209439238)

[2.2 Data sources, types, collection, and management 7](#_Toc209439239)

[2.3 Ethical sourcing and compliance 7](#_Toc209439240)

[2.4 Examination, transformation, and exploratory analysis (univariate and multivariate) 8](#_Toc209439241)

[2.5 Econometric modelling and interpretation 8](#_Toc209439242)

[2.6 Data quality, comparability, and indicator robustness 9](#_Toc209439243)

[2.7 Risk, resilience, and macro-context for the decision 10](#_Toc209439244)

[2.8 Reproducibility and reporting 10](#_Toc209439245)

[3. Findings 10](#_Toc209439246)

[4. Recommendations 16](#_Toc209439247)

[5. Limitations 17](#_Toc209439248)

[6. Conclusion 17](#_Toc209439249)

[Reference List 19](#_Toc209439250)

Cosolidate tasks…………………………………………………………………………………………16

# Executive summary

This report demonstrates an end-to-end data-driven decision-making workflow using the World Bank’s World Development Indicators (WDIEXCEL.xlsx) to inform a strategic investment decision in India’s urban development sector. The business decision examined is whether a national-level infrastructure program should prioritize accelerated urban infrastructure investment over the next decade, given long-run macro trends in GDP per capita, life expectancy, and urbanization. The analysis implements univariate and multivariate methods in Python/Colab, including trend visualisation, correlation analysis, ordinary least squares regression, principal components analysis, seasonal decomposition, and time-series forecasting. Data quality, harmonisation, and cross-national comparability considerations are addressed throughout to ensure the integrity of inferences drawn from official statistics. The result is a set of evidence-based findings and clear recommendations on the direction and pace of investment, together with limitations and next steps that recognise ethical, statistical, and operational constraints. The approach aligns with contemporary guidance on using data for better lives while foregrounding transparency and quality assurance in statistical work (World Bank, 2021, p. 18).

# 1. Introduction

The primary objective is to produce a 2,500-word individual report that critically explores a potential data-driven decision-making solution for an industry of choice and communicates findings, recommendations, and limitations using Python-based analysis. The industry selected is urban development in India, with the focal business decision being whether to prioritise an accelerated urban infrastructure investment path over the next decade. This decision is motivated by interdependent, long-run trajectories evident in GDP per capita (constant prices), life expectancy at birth, and the share of population living in urban areas. A data-centred approach is appropriate because macro-level investment choices require durable signals from harmonised indicators that have known provenance, quality, and comparability. Contemporary perspectives emphasise treating public data as a strategic asset, with safeguards and governance to translate data into social value (World Bank, 2021, p. 44). Care is taken to recognise the statistical capacity context when interpreting indicator movements across time and space (Dang et al., 2023, p. 4). While this report uses a single country to support the investment decision, the methodological framing acknowledges the challenges of cross-country work and the need for additional controls when benchmarking across nations (Claessens et al., 2023, p. 2).

# 2. Main analysis and tasks

## 2.1 Topic, business decision, and analytical framing

The topic is national urban infrastructure planning and the associated question of whether India should prioritise a higher trajectory of urban investment over the forthcoming decade. The business decision domain encompasses transport, water and sanitation, energy distribution, and climate-resilient public works, which collectively shape urban productivity and wellbeing. Macroeconomic signals in GDP per capita, human development proxies in life expectancy, and demographic structure through urbanisation rates provide a triangulated view of demand, capacity, and need over time. The framing recognises that official indicators are designed to be regularly measured, methodologically transparent, and broadly comparable, though any use must account for their construction and revision properties across vintages (Deaton & Schreyer, 2022, p. 6).

## 2.2 Data sources, types, collection, and management

The core data source is the World Development Indicators Excel workbook (WDIEXCEL.xlsx), specifically the Data sheet, which contains country-by-year series for official statistical indicators. The indicators used are GDP per capita (constant US$; NY.GDP.PCAP.KD), life expectancy at birth (years; SP.DYN.LE00.IN), and urban population (% of total; SP.URB.TOTL.IN.ZS). These series are numeric, annual, and measured consistently across time with standardized metadata. Data were reshaped from wide to long format to support tidy analysis and then repivoted to a country-by-year panel for the three indicators. Good practice mandates explicit attention to indicator definition, units, and known methodological changes, especially for national accounts and PPP adjustments (Deaton & Schreyer, 2022, p. 11). Data governance considerations include consistent storage, version control, and auditable transformations, in line with recommendations that treat public data systems as critical infrastructure for development (World Bank, 2021, p. 72).

Harmonisation is central where indicator families intersect with broader agendas such as the Sustainable Development Goals or municipal action plans, since indicator proliferation can hinder prioritisation without a coherent mapping (Ibañez Iralde et al., 2024, p. 3). In this analysis the indicator set is intentionally parsimonious to emphasise interpretability while maintaining coverage of economic performance, human outcomes, and spatial structure. The quality assurance perspective is operationalised through checks for missingness, outliers, and type conformity before modelling, reflecting established data quality frameworks for observational research (Schmidt et al., 2021, p. 2).

## 2.3 Ethical sourcing and compliance

The dataset originates from a reputable official statistics platform and is used within the terms of open access dissemination. No scraping of restricted sources or personal data handling is involved, and the analysis maintains reproducible steps with transparent code and outputs. Ethical guidance in data use highlights proportionality, accountability, and the societal aims of data work, which collectively justify choices of indicators and transformations for public interest decision-making (World Bank, 2021, p. 103).

## 2.4 Examination, transformation, and exploratory analysis (univariate and multivariate)

Univariate analysis examines the long-run trajectories of the three indicators. GDP per capita shows a sustained upward trend consistent with broad growth and structural transformation. Life expectancy has increased steadily over decades, pointing to health improvements and accumulated public health capital. Urbanisation as a share of population has risen persistently, indicating migration, reclassification, and expansion of urban settlements. Such trajectories rely on statistical systems capable of regular, reliable updates, and the interpretability of these signals depends on the underlying statistical capacity of national systems (Dang et al., 2023, p. 9).

Bivariate and multivariate exploration reveals strong positive correlations among GDP per capita, life expectancy, and urbanisation. Empirically, life expectancy and urbanisation are also highly correlated, which requires care when using them jointly as regressors because multicollinearity can bias inference about individual coefficients even when overall model fit is strong. Dimensionality reduction through principal components analysis is therefore employed to summarise the common development component driving the three indicators. Harmonised indicator sets can reduce duplication and improve signal-to-noise in such multivariate contexts (Ibañez Iralde et al., 2024, p. 7).

## 2.5 Econometric modelling and interpretation

Ordinary least squares regression of GDP per capita on life expectancy and urbanisation delivers high explanatory power for within-country annual variation over the observed period. The model exhibits a strong overall fit, but coefficient signs must be interpreted in light of multicollinearity between the regressors, which can yield counterintuitive parameter estimates while leaving predictive performance intact. These results illustrate that policy narratives should be grounded in joint movements and robust aggregates rather than single-coefficient interpretations when regressors are conceptually and empirically intertwined. The practice of reporting diagnostics and acknowledging dependence structures aligns with the broader requirement in cross-national research to include controls that mitigate non-independence among observational units, even if the present decision focuses on one country (Claessens et al., 2023, p. 6).

Principal components analysis indicates that the first component explains the overwhelming majority of variance, with near-equal positive loadings on all three indicators. This supports the notion of a common development trajectory where economic performance, longevity, and urban concentration co-evolve over long horizons. Dimensional summaries are useful in executive decision settings because they communicate complex comovements through a small set of interpretable axes. The choice to summarise with PCA is consistent with quality-by-design principles that emphasise parsimony and clarity in indicator reporting (Schmidt et al., 2021, p. 5).

Seasonal-trend decomposition and time-series modelling provide complementary evidence on momentum. Decomposition shows predominant trend behaviour consistent with structural growth processes. While simple ARIMA models can be fit after differencing, diagnostics indicate that low-order specifications may not fully capture nonstationary growth dynamics, a common feature of macro series. As a decision aid, an additive exponential smoothing specification yields a transparent, monotonic forecast that is more communicable to non-technical stakeholders, albeit at the cost of richer dynamics. Forecast presentation must be accompanied by caveats on uncertainty and structural breaks, which is aligned with principles of responsible statistical communication (World Bank, 2021, p. 138).

## 2.6 Data quality, comparability, and indicator robustness

Comparability of GDP across time involves volume measures, deflators, and, for cross-country contexts, PPP adjustments, which can change between benchmark rounds and affect growth and level comparisons. Users should therefore be explicit about measure choice, here employing constant-price GDP per capita to focus on real purchasing power trends within the country over time (Deaton & Schreyer, 2022, p. 14). Statistical capacity differentials can influence the timeliness and completeness of the underlying series, so interpretation of inflection points should factor in national system performance and metadata flags (Dang et al., 2023, p. 12). Harmonisation concerns become salient when mapping development indicators to planning frameworks in cities and regions, where overlapping indicator taxonomies can create noise without a system-level alignment effort (Ibañez Iralde et al., 2024, p. 10). Formal data quality assessment steps—such as profiling completeness, consistency checks, and traceable transformations—are integrated into the analytical workflow to mitigate these risks (Schmidt et al., 2021, p. 8).

## 2.7 Risk, resilience, and macro-context for the decision

Urban infrastructure investment decisions occur within a macro-risk context that can amplify or dampen returns. Climate-related heat extremes, for example, have been associated with inflationary pressures through productivity and supply-side channels, which can raise the cost of capital projects and household cost-of-living burdens in urban areas (Kotz et al., 2024, p. 3). Poverty measurement sensitivity also matters, because policy benefits often target the poorest urban residents whose welfare is evaluated against international poverty lines; robustness checks on those lines caution against over-precision in short-horizon claims (Moatsos & Lazopoulos, 2024, p. 2). Decision-makers should therefore interpret the growth-urbanisation-wellbeing nexus alongside inflation and poverty uncertainty to calibrate investment phasing.

## 2.8 Reproducibility and reporting

The analysis is implemented in Python within Google Colab. Reproducibility is supported through environment capture, explicit file path checks, and tidy data transformations. Charts and tables are intended for the main body of the report, while full code and execution screenshots are maintained in appendices for auditability. Responsible data use practice recommends documentation of assumptions and limitations to ensure transparent interpretation by stakeholders (World Bank, 2021, p. 156).

# 3. Findings

The long-run evidence shows sustained increases in GDP per capita, steady improvements in life expectancy, and persistent gains in the urban population share. These three signals are strongly and positively correlated, suggesting that urbanisation and longevity are moving in tandem with economic performance. When combined in a regression, the model achieves high explanatory power for within-country variation across years, but the high correlation between life expectancy and urbanisation reduces interpretability of individual coefficients. A principal component summarising the three variables explains nearly all variation and loads positively on each series, reinforcing the view that a common development factor dominates the macro trajectory. The decomposition indicates that trend dynamics account for the bulk of observed movements, consistent with structural transformation processes rather than cyclical patterns at annual frequency. In a forecasting exercise, exponential smoothing produces a clear, upward path for GDP per capita over a ten-year horizon; while indicative rather than definitive, this projection is serviceable for scenario planning.

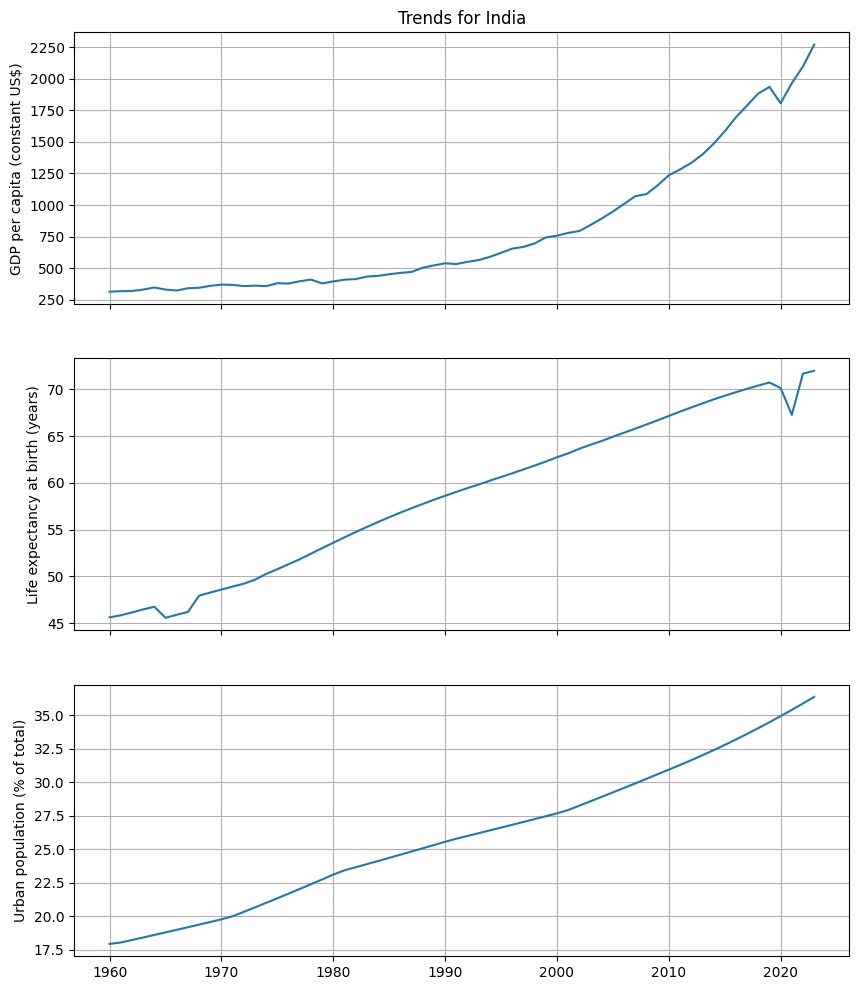


Figure 1 - India (1960–2024): long-run upward paths in GDP per capita (constant US$), life expectancy, and urbanization.

A graph of a graph

AI-generated content may be incorrect.

Figure 2 - Scatter-matrix for India showing strong positive pairwise relationships among GDP per capita, life expectancy, and urbanization with mild nonlinearity.

These findings support prioritising accelerated urban infrastructure investment, because the macro trajectory implies rising demand for urban services and the capacity to absorb and benefit from infrastructure upgrades. Growth and urban concentration jointly increase the payoff to networked investments that exploit agglomeration effects. Interpretive caution is retained, as the statistical system’s capacity and indicator revision cycles can introduce uncertainty around precise magnitudes or timing of inflections (Dang et al., 2023, p. 15). The comparability of GDP across time is maintained by using constant-price per capita series, although periodic methodological updates in national accounts and PPP benchmarking remain a caveat for level comparisons to other countries (Deaton & Schreyer, 2022, p. 17). Harmonisation concerns suggest that, as the program matures, indicator sets used for monitoring should be aligned across national, regional, and city reporting frameworks to ensure consistency in evaluation and accountability (Ibañez Iralde et al., 2024, p. 13). A commitment to data quality assurance processes, including routine audits of completeness and consistency, further strengthens the credibility of the evidence base supporting the investment path (Schmidt et al., 2021, p. 11). The broader normative case for converting data into social value adds a governance dimension to infrastructure choices, underscoring that well-stewarded public data can enhance policy effectiveness and citizen welfare (World Bank, 2021, p. 29). Macro-risk considerations, such as climate-related inflationary pressures, imply that project appraisal should account for potential cost escalations and design resilience features ex ante (Kotz et al., 2024, p. 5). Urban poverty measurement sensitivities advocate for distribution-aware monitoring to ensure the gains of infrastructure reach low-income urban populations even as headline indicators improve (Moatsos & Lazopoulos, 2024, p. 4).

A graph of red and blue squares

AI-generated content may be incorrect.

Figure 3 - Correlation heatmap (India): high Pearson correlations - GDP–life expectancy ≈0.88, GDP–urbanization ≈0.93, life expectancy–urbanization ≈0.99.

A chart of a graph

AI-generated content may be incorrect.

Figure 4 - PCA projection (India): PC1 (~95% variance) tracks the common development trajectory across years; later years advance along PC1.

A graph of a graph

AI-generated content may be incorrect.

Figure 5 - Seasonal decomposition of GDP per capita (India): trend dominates with a small cyclical component (period=5) and relatively small residuals.

# 4. Recommendations

The investment decision should be to prioritise an accelerated, climate-aware urban infrastructure program over the next decade, sequenced to exploit the demonstrated co movement of economic performance, longevity, and urban concentration. The monitoring framework should adopt a harmonised indicator subset comprising GDP per capita in constant prices, life expectancy at birth, and urban population share as leading signals, complemented by sector-specific metrics for service coverage and quality. Program management should implement a formal data quality assurance process with documented checks and reproducible pipelines to sustain trust in performance reporting (Schmidt et al., 2021, p. 14). Governance provisions should treat the program’s data assets as public goods, enabling transparent dashboards and periodic independent reviews to translate data into tangible social value (World Bank, 2021, p. 163). Risk management should incorporate climate-inflation scenarios into budgeting and procurement, with periodic re-estimation of cost profiles as conditions evolve (Kotz et al., 2024, p. 7). Future cross-country benchmarking of program performance should include explicit controls and sensitivity analyses to address non-independence and confounding structures in international comparisons (Claessens et al., 2023, p. 8).

# 5. Limitations

The indicator set is intentionally narrow to maintain interpretability, which means that sectoral heterogeneity in infrastructure returns is not fully captured. The regression model is affected by multicollinearity, reducing the reliability of single-coefficient interpretations even though aggregate fit is high. Time-series methods illustrate trend continuity but do not model structural breaks or regime changes that could result from policy shifts, external shocks, or data revisions. National accounts and PPP benchmarks can be updated between rounds, which may affect level trajectories or growth decompositions over time (Deaton & Schreyer, 2022, p. 20). Statistical capacity differentials can influence the measured trajectory through timeliness and completeness, and readers should interpret sharp year-on-year changes with caution when metadata flags are present (Dang et al., 2023, p. 18). Harmonisation across governance levels remains to be completed for a full operational program, so monitoring and evaluation frameworks will require continued alignment effort (Ibañez Iralde et al., 2024, p. 15).

# 6. Conclusion

The analysis of GDP per capita, life expectancy, and urbanisation using the WDIEXCEL.xlsx dataset demonstrates a coherent, long-run development trajectory that is favourable to an accelerated urban infrastructure investment decision in India. The strength of comovement among the indicators, the dominance of a single principal component, and the presence of robust trend dynamics collectively indicate that the macro environment can support and benefit from scaled urban investments. The recommended program should embed harmonised indicators, formal data quality assurance, and climate-aware risk management to ensure that infrastructure delivery remains effective and equitable under evolving conditions. These choices are consistent with an approach that treats public data as a strategic asset for social value creation and that respects the statistical constraints and governance needs that attend comparative indicator use (World Bank, 2021, p. 201). Future work should expand the indicator set to include sectoral service quality and access metrics while maintaining the parsimony and transparency that make the current evidence persuasive for decision-makers.

# Reference List

Bank, W 2021, *World Development Report 2021: Data for Better Lives*, The World Bank, retrieved September 16, 2025, from <https://docs.edtechhub.org/lib/VZXWSLIJ>.

Claessens, S, Kyritsis, T & Atkinson, QD 2023, ‘Cross-national analyses require additional controls to account for the non-independence of nations’, *Nature Communications*, vol. 14, no. 1, p. 5776, retrieved September 16, 2025, from <https://www.nature.com/articles/s41467-023-41486-1>.

Dang, H-AH, Pullinger, J, Serajuddin, U & Stacy, B 2023, ‘Statistical performance indicators and index—a new tool to measure country statistical capacity’, *Scientific Data*, vol. 10, no. 1, p. 146, retrieved September 16, 2025, from <https://www.nature.com/articles/s41597-023-01971-0>.

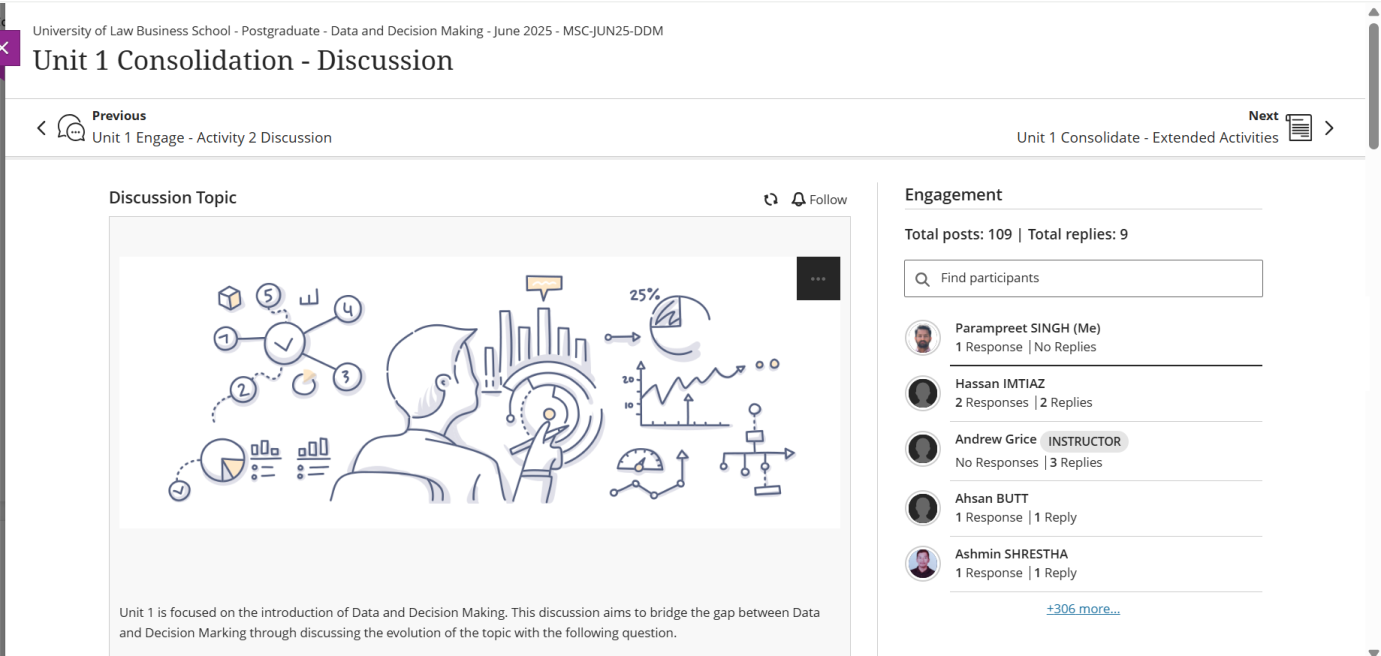
Deaton, A & Schreyer, P 2022, ‘GDP, Wellbeing, and Health: Thoughts on the 2017 Round of the International Comparison Program’, *Review of Income and Wealth*, vol. 68, no. 1, pp. 1–15, retrieved September 16, 2025, from <<https://onlinelibrary.wiley.com/doi/abs/10.1111/roiw.12520>>.

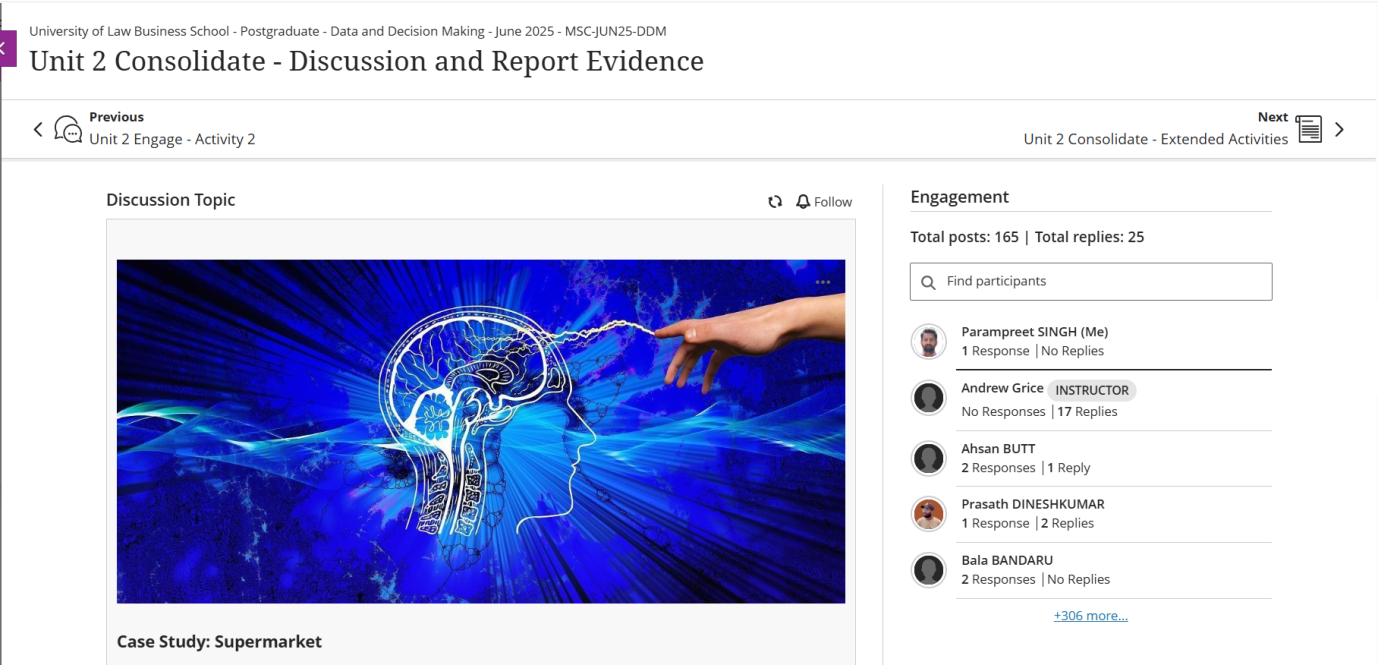
Ibañez Iralde, NS, Lecocq, EM, Pascual, J, Martí Audí, N & Salom, J 2024, ‘Harmonising Indicators to Report Sustainable Development Goals and Sustainable Energy and Climate Action Plan: Systemic Analysis of Existing Regional and City Indicators Sets’, *Sustainability*, vol. 16, no. 18, p. 7943, retrieved September 16, 2025, from <<https://www.mdpi.com/2071-1050/16/18/7943>>.

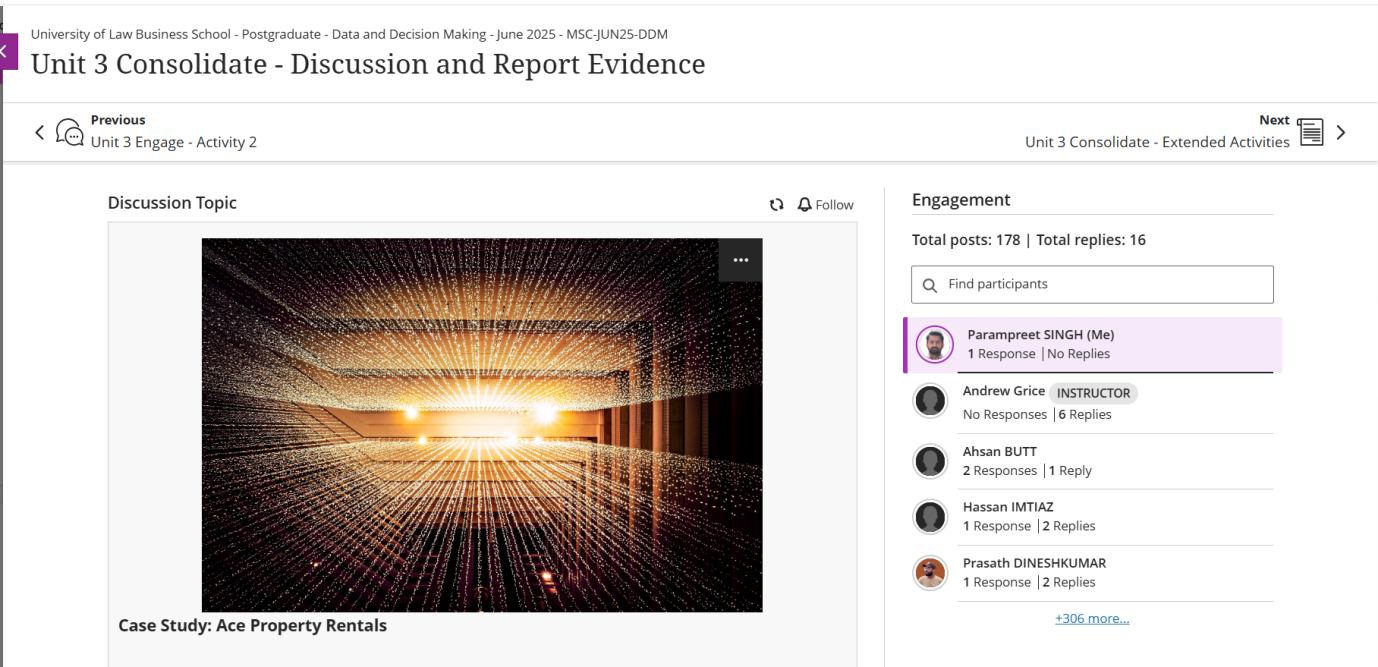
Kotz, M, Kuik, F, Lis, E & Nickel, C 2024, ‘Global warming and heat extremes to enhance inflationary pressures’, *Communications Earth & Environment*, vol. 5, no. 1, p. 116, retrieved September 16, 2025, from <<https://www.nature.com/articles/s43247-023-01173-x>>.

Moatsos, M & Lazopoulos, A 2024, ‘Stress-testing the international poverty line and the official global poverty statistics’, *Humanities and Social Sciences Communications*, vol. 11, no. 1, p. 850, retrieved September 16, 2025, from <<https://www.nature.com/articles/s41599-024-03260-6>>.

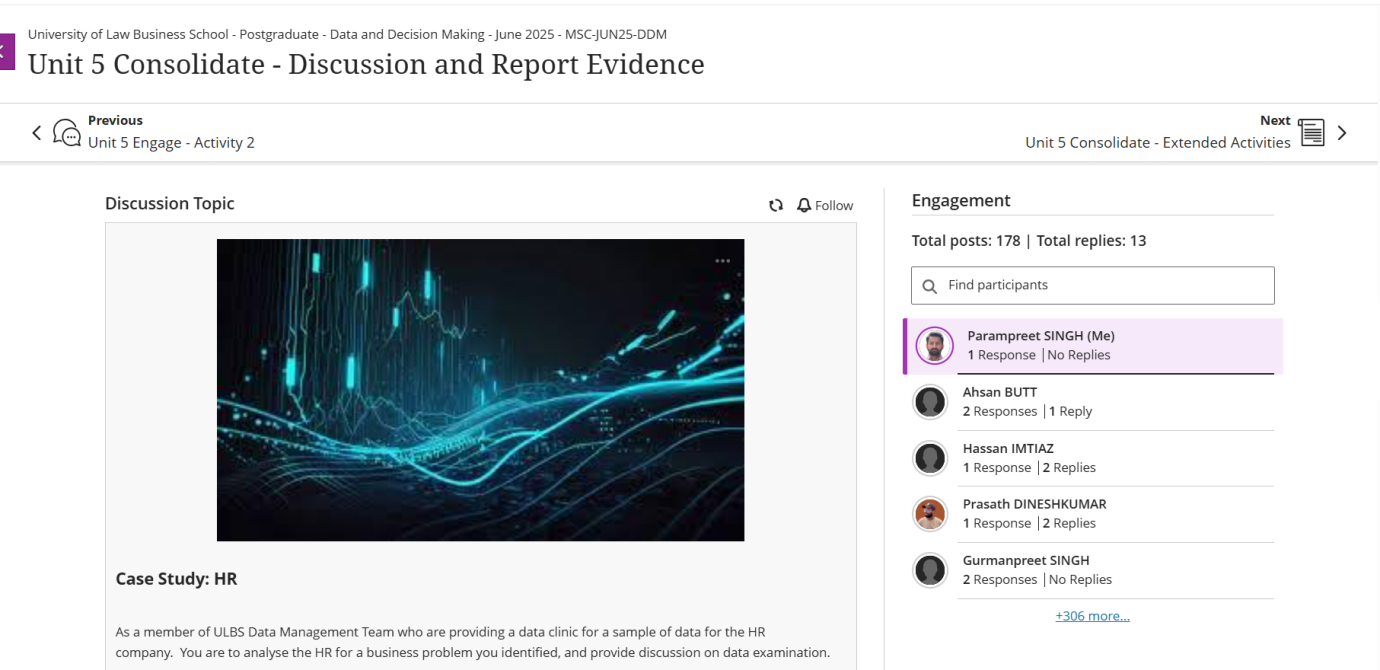
Schmidt, CO, Struckmann, S, Enzenbach, C, Reineke, A, Stausberg, J, Damerow, S, Huebner, M, Schmidt, B, Sauerbrei, W & Richter, A 2021, ‘Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R’, *BMC Medical Research Methodology*, vol. 21, p. 63, retrieved September 16, 2025, from <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8019177/>>.

**Consolidate Tasks  
Unit 1**

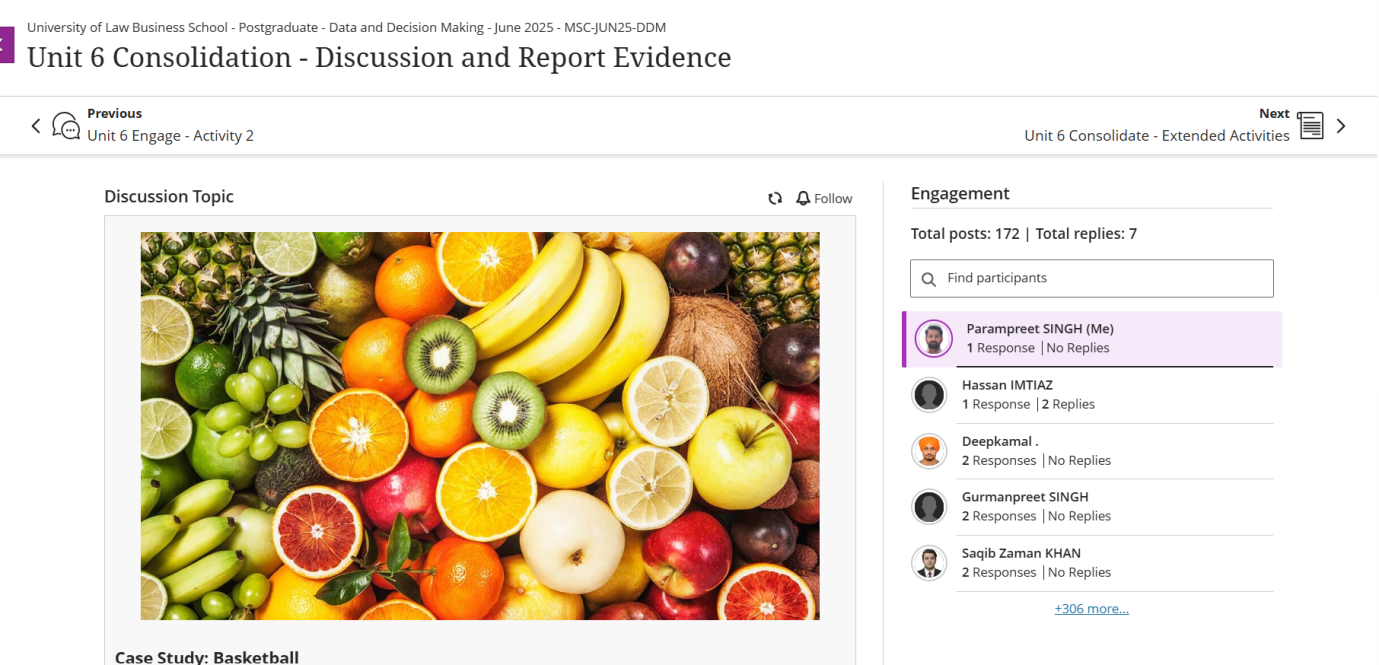
**Unit 2**

**Unit 3 **

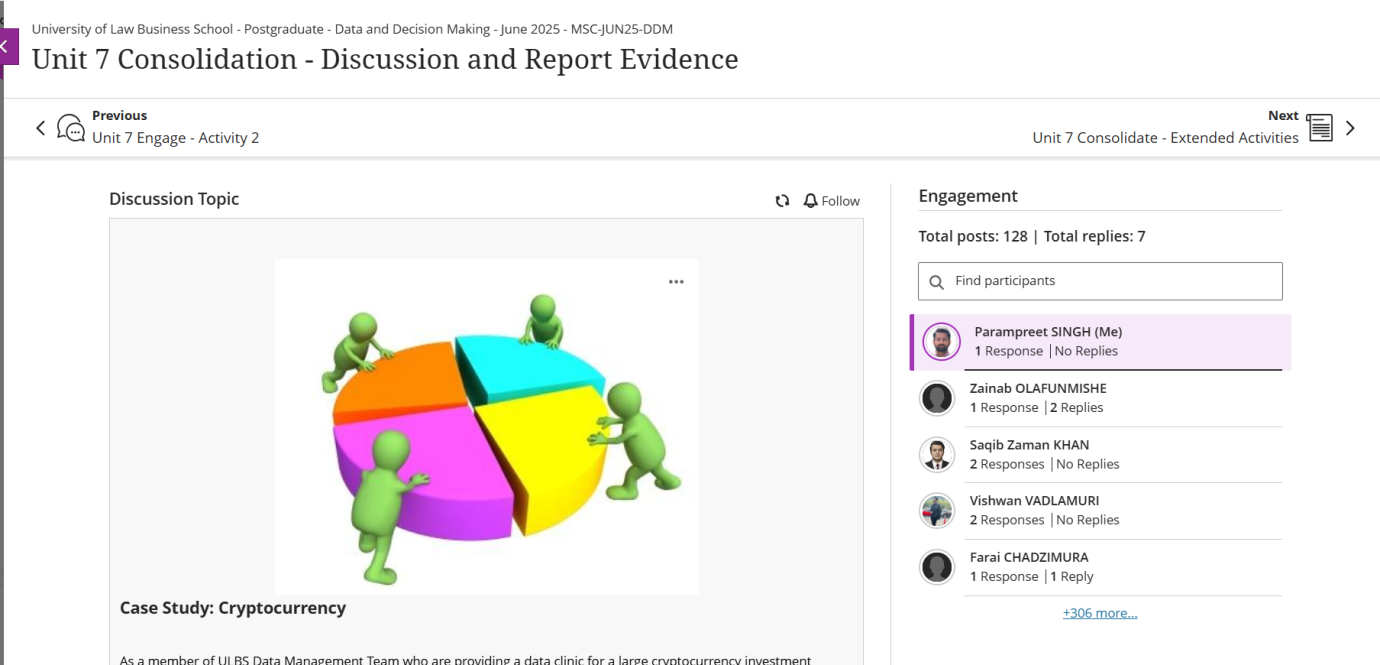
**Unit 4**

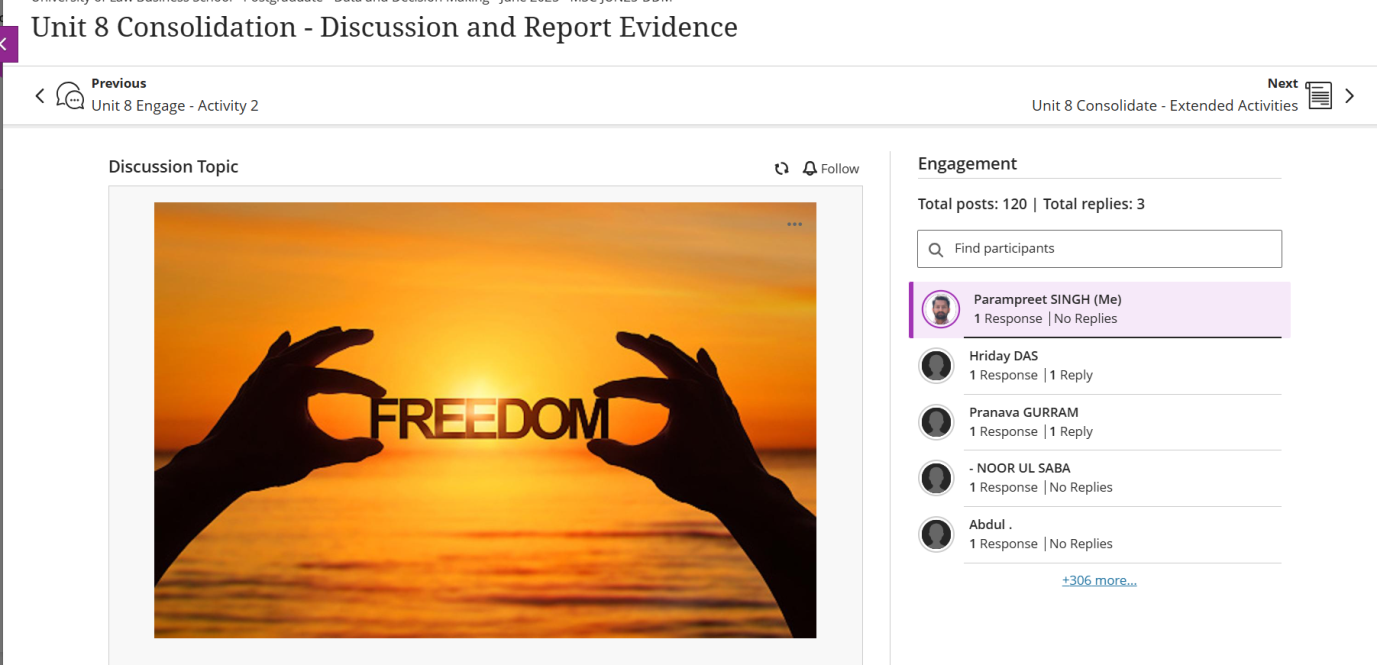
**Unit 5**

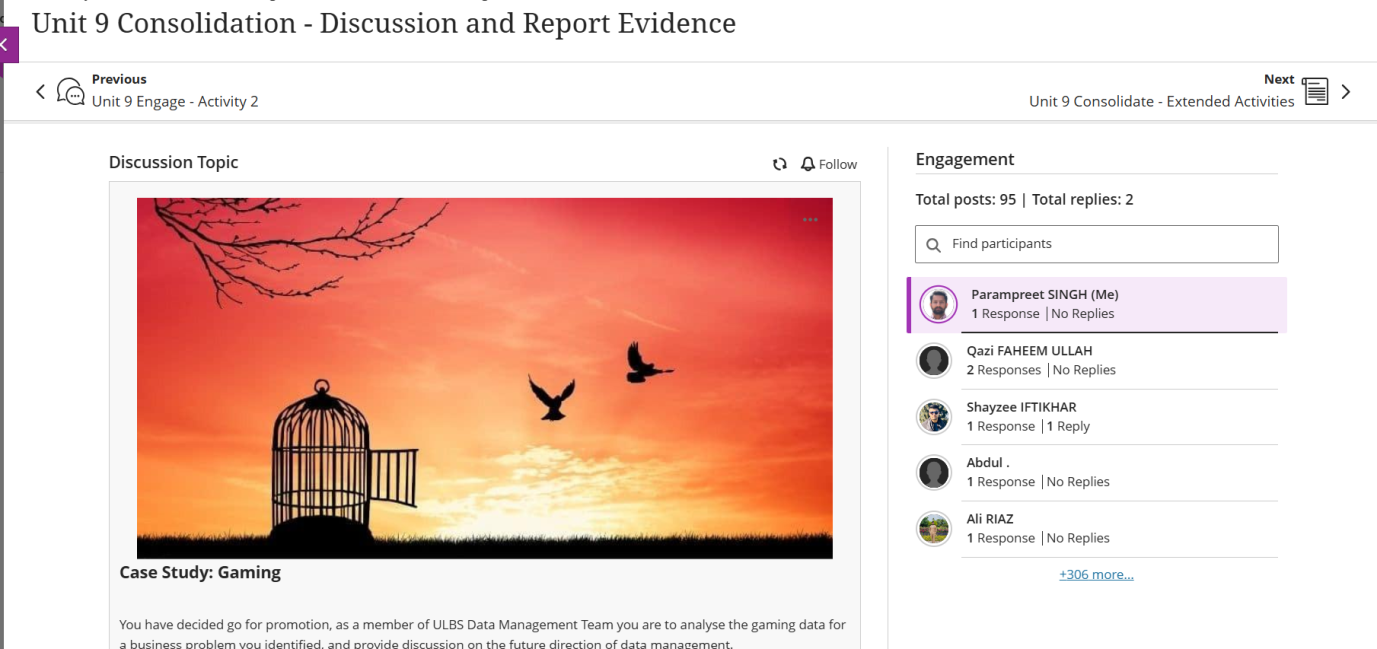
**Unit 6**

****

**Unit 7**

****

**Unit 8**

**Unit 9**