VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

MGT415

Introduction to Business Analytics

MGT415 by VU Learning HUB

Prepared for VU students • Free study resources

https://vulearninghub.xyz

MGT415

Introduction to Business Analytics

STUDY HANDOUT

A complete topic-wise study handout for MGT415 Introduction to Business Analytics, prepared from business analytics lecture series. Covers what analytics is, data and databases, descriptive statistics, data visualisation, probability foundations, regression, forecasting and prescriptive models - with concepts, worked examples, Excel-oriented guidance and exam pointers.

VU Learning HUB | vulearninghub.xyz

Table of Contents

1. What Is Business Analytics
2. Why Analytics Matters for Business
3. Data: Types and Sources
4. Databases and Data Warehousing
5. Descriptive Statistics I: Centre and Spread
6. Descriptive Statistics II: Shape and Position
7. Data Visualisation Principles
8. Probability Foundations
9. Probability Distributions
10. Sampling and Estimation
11. Hypothesis Testing
12. Correlation and Association
13. Simple Linear Regression
14. Multiple Regression
15. Forecasting I: Time Series Basics
16. Forecasting II: Accuracy and Practice
17. Prescriptive Analytics: Optimisation
18. Simulation and What-If Analysis
19. Analytics in Marketing
20. Analytics in Operations and Finance
21. Data Ethics and Privacy
22. Communicating Results and Storytelling
23. Revision: Midterm Practice Set
24. Revision: Final-Term Practice Set

Lecture 1: What Is Business Analytics

Key terms: business analytics, data-driven decisions, DIKW

Business analytics is the use of data, statistical analysis and modelling to support business decisions. It turns raw data into insight: what happened, why it happened, what will happen, and what should be done. The DIKW hierarchy captures the journey - Data (raw facts) -> Information (organised data) -> Knowledge (patterns) -> Wisdom (decisions). Analytics is the engine of that journey.

- Analytics answers: descriptive (what happened), diagnostic (why), predictive (what next), prescriptive (what to do).
- Data-driven firms decide faster and with less bias - the competitive edge.
- Analytics complements, not replaces, managerial judgement.

EXAM POINTER: Explain the four types of analytics with a retail example for each.

Lecture 2: Why Analytics Matters for Business

Key terms: competitive advantage, evidence-based management

Businesses drown in data - sales transactions, web clicks, sensor readings. Analytics converts this flood into pricing decisions, inventory plans, customer targeting and risk control. Evidence-based management replaces gut feel with tested insight: banks score credit, retailers optimise shelves, airlines price seats dynamically - all analytics at work.

- Revenue: dynamic pricing, cross-selling, churn reduction.
- Cost: demand forecasting cuts inventory; predictive maintenance cuts downtime.
- Risk: fraud detection, credit scoring, compliance monitoring.

Lecture 3: Data: Types and Sources

Key terms: structured, unstructured, internal, external data

Data comes structured (tables: rows and columns - sales records) or unstructured (text, images, video, social posts). Sources: internal (ERP, CRM, POS systems) and external (market data, government statistics, web scraping, surveys). Quality dimensions: accuracy, completeness, consistency, timeliness - garbage in, garbage out remains the iron law.

- Structured data fits spreadsheets and databases; unstructured needs special tools.
- Big data's 3 Vs: volume, velocity, variety (plus veracity).
- Always audit data quality before analysis - check for missing and duplicate records.

Lecture 4: Databases and Data Warehousing

Key terms: database, SQL, data warehouse, ETL

Operational data lives in databases (relational tables queried with SQL); analytical data is consolidated in data warehouses - subject-oriented, historical stores built via ETL (extract, transform, load). A data mart is a departmental slice. This separation exists because transaction systems optimise for speed of recording, warehouses for speed of questioning.

- SQL basics for analysts: SELECT, WHERE, GROUP BY, JOIN.
- ETL cleans and standardises data from many sources into one truth.
- Star schema: fact tables (numbers) joined to dimension tables (context).

Lecture 5: Descriptive Statistics I: Centre and Spread

Key terms: mean, median, mode, variance, standard deviation

Descriptive statistics summarise data. Central tendency: mean (average), median (middle value), mode (most frequent). Dispersion: range, variance, standard deviation (typical distance from the mean). The median resists outliers - house prices are reported as medians because a few mansions distort the mean. Standard deviation drives risk measurement in finance.

- Skewed data: report median; symmetric data: mean is fine.
- Standard deviation: ~68% of normal data lies within one SD of the mean.
- Excel: AVERAGE, MEDIAN, MODE, STDEV, VAR functions.

Lecture 6: Descriptive Statistics II: Shape and Position

Key terms: skewness, kurtosis, percentiles, quartiles, boxplot

Shape: skewness (asymmetry - income data skews right), kurtosis (tailedness). Position: percentiles and quartiles locate values - the median is the 50th percentile. The boxplot displays quartiles, median and outliers in one picture, making it the analyst's quickest distribution check. Z-scores standardise values: how many SDs from the mean.

- Right skew: mean > median (a few large values pull the mean up).
- IQR = Q3 - Q1; points beyond 1.5xIQR are flagged as outliers.
- Z-score beyond +/-3 deserves investigation.

EXAM POINTER: Given a salary dataset, compute mean, median, SD and interpret the skewness.

Lecture 7: Data Visualisation Principles

Key terms: visualisation, charts, dashboards

A chart should reveal truth, not decorate. Match chart to purpose: trends over time -> line chart; comparisons -> bar chart; composition -> pie (sparingly) or stacked bar; distribution -> histogram; two-variable relationships -> scatter plot. Principles: start bars at zero, label directly, avoid 3D and chartjunk, one message per chart. Dashboards assemble key charts for monitoring.

- The right chart answers the question faster than any table.
- Dual axes mislead - use with care and clear labelling.
- Colour: highlight the message, not the rainbow.

Lecture 8: Probability Foundations

Key terms: probability, conditional probability, Bayes

Probability quantifies uncertainty on a 0-1 scale. Rules: P(A or B) = P(A) + P(B) - P(A and B); independent events multiply: P(A and B) = P(A)P(B). Conditional probability P(A|B) updates beliefs with evidence; Bayes' theorem formalises it - the logic behind spam filters, medical testing interpretation, and customer targeting models.

- Base-rate neglect: a 99% accurate test still misleads for rare diseases - Bayes shows why.
- Expected value: probability x payoff - the basis of business bets.
- Independence means one event doesn't change the other's probability.

Lecture 9: Probability Distributions

Key terms: normal, binomial, Poisson, distributions

Distributions describe patterns of randomness. Binomial: counts of successes in n trials (defective items). Poisson: events per interval (customers per hour). Normal (bell curve): the workhorse - heights, measurement errors, averages; the Central Limit Theorem says sample means become normal as samples grow, which is why normal-based inference works so widely.

- Normal: defined by mean and SD; 95% within ~2 SDs.
- Poisson for arrivals/queues; binomial for yes/no counts.
- Check distribution shape before choosing a model.

Lecture 10: Sampling and Estimation

Key terms: sampling, confidence interval, margin of error

We rarely observe whole populations, so we sample - randomly, to avoid bias. A point estimate (sample mean) guesses the population value; a confidence interval gives a range: 'we are 95%

confident the true mean lies between X and Y.' Larger samples narrow intervals but cost more - the classic trade-off. Biased sampling (convenience, self-selection) invalidates everything.

- Random sampling is the foundation - without it, inference collapses.
- Margin of error shrinks with the square root of sample size: 4x data halves error.
- 95% confidence does not mean 95% probability the parameter is inside.

EXAM POINTER: Explain confidence intervals and why random sampling is non-negotiable.

Lecture 11: Hypothesis Testing

Key terms: hypothesis test, p-value, significance, Type I/II errors

Hypothesis testing asks: could this result be mere chance? Set H0 (no effect) vs H1 (effect), compute a test statistic, and read the p-value - the probability of data this extreme if H0 were true. $p < 0.05$ rejects H0 (statistically significant). Errors: Type I (false alarm, probability alpha) and Type II (missed effect, beta). Business use: A/B testing of websites, ads, and offers.

- A/B test: split traffic, measure conversion, test the difference.
- Significant does not mean important - check effect size too.
- Multiple testing inflates false positives; adjust or pre-register.

Lecture 12: Correlation and Association

Key terms: correlation, causation, scatter plot

Correlation (r, from -1 to +1) measures linear association: advertising spend vs sales often correlate strongly. But correlation is not causation - ice-cream sales correlate with drowning (both rise in summer). Confounders, reverse causality and coincidence all fake causation. Experiments (randomised trials) establish cause; observational data only suggests it.

- r near +1/-1: strong linear link; near 0: none (linear).
- Always plot: Anscombe's quartet shows identical correlations, wildly different data.
- Spurious correlations are everywhere in big data - domain sense filters them.

Lecture 13: Simple Linear Regression

Key terms: regression, least squares, R-squared

Regression models Y as a function of X: $Y = a + bX + \text{error}$. Least squares finds the line minimising squared errors; b is the slope (effect of one more unit of X), a the intercept. R-squared measures fit: the fraction of Y's variation explained (0.85 = 85%). Use: predicting sales from ad spend, house prices from area. Check residuals - patterns there mean the model is wrong.

- Interpretation: b = expected change in Y per unit of X.
- R-squared high but residuals patterned = misleading model.
- Never extrapolate far beyond the data range.

Lecture 14: Multiple Regression

Key terms: multiple regression, multicollinearity, dummy variables

Real outcomes have many drivers: sales = f(price, advertising, season, competitor action). Multiple regression estimates each factor's effect holding others constant. Watch for multicollinearity (predictors correlated with each other - inflates uncertainty) and use dummy variables for categories (season, region). Adjusted R-squared penalises adding useless predictors.

- Ceteris paribus interpretation: effect of X1 with other Xs fixed.
- VIF > 10 signals harmful multicollinearity - drop or combine predictors.
- Dummy trap: with k categories use k-1 dummies.

Lecture 15: Forecasting I: Time Series Basics

Key terms: time series, trend, seasonality, moving average

Time series data has structure: trend (long-run direction), seasonality (regular cycles - Eid sales spikes), and noise. Decomposition separates them. Simple methods: moving averages smooth noise; exponential smoothing weights recent observations more. Forecasts need the right granularity - daily for staffing, monthly for budgeting.

- Plot first: trend and seasonality are visible before any model.
- Seasonal businesses must deseasonalise before judging performance.
- Naive forecast (next = last) is the benchmark to beat.

EXAM POINTER: Decompose a retail sales series into trend, seasonal and residual components.

Lecture 16: Forecasting II: Accuracy and Practice

Key terms: forecast accuracy, MAPE, judgmental forecasting

Measure accuracy: MAE (average error), MAPE (percentage error - intuitive for business). Compare models on holdout data - fit on history, test on the most recent period. Combine statistical forecasts with judgement (promotions, disruptions the data can't see). Honest practice: report uncertainty bands, update as new data arrives, and track bias (chronic over/under-forecasting).

- MAPE: 'we are off by 8% on average' - managers understand it.
- Holdout testing prevents overfitting to history.
- Document assumptions - forecasts are only as good as their premises.

Lecture 17: Prescriptive Analytics: Optimisation

Key terms: optimisation, linear programming, Solver

Prescriptive analytics answers 'what should we do?' Optimisation finds the best decision under constraints: linear programming allocates limited resources (production mix, diet problems, transport) to maximise profit or minimise cost. Excel Solver implements it: define the objective cell, decision cells, and constraints - then solve. Shadow prices reveal which constraints bite hardest.

- LP ingredients: objective, decision variables, constraints - all linear.
- Solver's sensitivity report shows how stable the optimum is.
- Real problems add integer or nonlinear twists - start simple.

Lecture 18: Simulation and What-If Analysis

Key terms: simulation, Monte Carlo, scenario analysis

When uncertainty defeats formulas, simulate: Monte Carlo runs thousands of random scenarios through a model to map possible outcomes - project completion times, investment returns, inventory needs. Excel's data tables and Scenario Manager do simpler what-if analysis: change inputs, watch outputs. Decision trees structure sequential choices with probabilities at each branch.

- Monte Carlo: replace single guesses with distributions of inputs.
- Output is a distribution: P(profit > target), not a single number.
- Tornado charts rank which uncertainties matter most.

Lecture 19: Analytics in Marketing

Key terms: marketing analytics, segmentation, CLV, churn

Marketing analytics: segment customers (RFM: recency, frequency, monetary), target with models (propensity to buy), measure campaigns (lift, ROI), and fight churn (predict who leaves, intervene). Customer lifetime value (CLV) focuses spending on the customers worth keeping. Digital marketing made everything measurable - attribution models divide credit among touchpoints.

- RFM segmentation: simple, powerful, needs only transaction data.
- Churn models flag at-risk customers before they leave.

- Test and learn: every campaign is an experiment.

Lecture 20: Analytics in Operations and Finance

Key terms: operations, inventory, credit scoring, fraud

Operations: EOQ balances ordering vs holding costs; safety stock covers demand uncertainty; queueing models staff service counters. Finance: credit scoring predicts default (logistic regression classics), fraud detection flags anomalies, portfolio models balance risk and return. Across both, dashboards monitor KPIs and alert on exceptions.

- EOQ: order quantity minimising total inventory cost.
- Credit scores: data-driven lending, monitored for bias and drift.
- Anomaly detection: unusual patterns trigger human review.

Lecture 21: Data Ethics and Privacy

Key terms: ethics, privacy, bias, GDPR

Analytics power demands responsibility: privacy (collect minimally, anonymise, secure), consent (tell people how data is used), and bias (models trained on biased history discriminate - hiring, lending). Regulations (GDPR-style laws) grant rights of access and erasure. Ethical analytics also means honesty: no cherry-picked charts, no hidden assumptions, uncertainty disclosed.

- Anonymisation is harder than it looks - re-identification attacks work.
- Audit models for disparate impact across groups.
- Document data lineage: where data came from and what was done to it.

EXAM POINTER: Discuss two ethical risks of business analytics and how firms should manage them.

Lecture 22: Communicating Results and Storytelling

Key terms: storytelling, communication, recommendations

Analysis without influence is wasted. Structure findings as a story: situation, complication, question, answer - then evidence. Lead with the recommendation, support with the three strongest facts, show one clear chart per point, and end with next steps. Know your audience: executives want implications and decisions, not methodology - keep the technical appendix for the appendix.

- BLUF: bottom line up front - recommendation first.
- One chart, one message; annotate the takeaway on the chart.

• Anticipate objections and address them before questions.

Lecture 23: Revision: Midterm Practice Set

1. Explain the four types of analytics with business examples. (10 marks)
2. A dataset of 200 shop sales: compute mean, median, SD; interpret skewness. (10 marks)
3. Design a dashboard for a retail store manager: choose 6 KPIs and chart types. (10 marks)
4. Explain Bayes' theorem with the medical-testing example. (5 marks)
5. What is a confidence interval? Why does random sampling matter? (10 marks)

Lecture 24: Revision: Final-Term Practice Set

1. Fit a simple regression of sales on advertising: interpret slope and R-squared. (10 marks)
2. Explain multicollinearity and how to detect and handle it. (10 marks)
3. Forecast next quarter's demand: outline the full process. (10 marks)
4. Formulate a product-mix linear programme for a factory. (10 marks)
5. Discuss ethical risks in analytics: privacy and bias. (5 marks)

End of Handout - Good luck with your exams!