RSU STUDY ABROAD

Big Data in Digital Business

Analysing digital-marketing data with an AI assistant you check by hand

Course code: IDB624 · **Level:** introductory (no technical background needed) · **Length:** 14 sessions, 45 contact hours, 14 November to 12 December 2026 · **Session:** 180 minutes; the three Saturday-afternoon labs run 240

Taught by Watthanasak Jeamwatthanachai, BEng, MSc, PhD, EMBA, C|CISO

Welcome

Every online shop now collects more data than any person can read. Each order, click, email and review becomes a row in a table, and a busy shop has millions of rows. So when a manager asks "how much did we really sell?" or "which customers should get the voucher?", a computer has to find the answer. More and more often, an AI assistant does that computer work. You describe what you want in plain English, the assistant writes the code, and a number appears.

This course teaches you to work that way without being fooled. The AI is fast, but it is sometimes wrong, and a wrong number looks exactly like a right one. So in every session you first work the method by hand on a small slice of the data: a **hand card** of five to twelve rows. Then you ask the AI to do the same steps, and you trust its answer on the full table only when its numbers match yours. You will never write code from an empty page. You will write clear steps, check results, catch mistakes, and turn checked numbers into a marketing decision you can defend.

The course follows one real company, an online gift seller in the UK, through the questions its marketing team would ask. How much did we really sell? Who are our best customers? Are we keeping them? Then it brings in other kinds of data: an email experiment, website visits and customer reviews. It ends with the rules on personal data and with your own final project: one marketing decision, analysed with the AI, checked by hand, and defended in front of the class.

Who this course is for

This course is for master's students in Rangsit University's Digital Business (MIDB) programme, and for any student who registers for the elective. Most students come from marketing or business administration. Most have never written code or opened a notebook. Some use spreadsheets every day at work, and some have never used one. All of that is fine, because the course starts from zero.

The course is taught in English, and most of the class works in English as a second language. So the slides and notebooks use short sentences and plain words, and every new term is explained the first time it appears.

What you will be able to do

- Compute revenue, average order value, RFM scores, retention, uplift against a control group, basket lift and a churn score's hit rate on a hand-sized case, and use the result to accept or reject an AI assistant's full-scale answer
- Formulate labelled plain-English analysis steps that an AI assistant in Google Colab turns into a full-scale analysis, and verify the run by reconciling its row counts, its totals and one pulled-back customer
- Diagnose flaws in an AI-generated marketing analysis, such as double counts, silent drops, partial periods, leakage and invented numbers, and state the fix for each
- Analyze web-session, funnel and customer-review data to locate where visitors and customers are lost, and justify where marketing effort should go
- Defend a marketing decision in a two-page memo, a presentation and an oral defence, using checked numbers, the strength of the evidence, a test that would settle it and what would change the call
- Assess a proposed use of customer data against Thailand's PDPA, the GDPR where it applies and an AI tool's data terms, and propose a compliant alternative

What you need first

- None. This course starts from first principles. You need a laptop and a Google account (see below), but no knowledge of statistics, spreadsheets or programming.

Before you start

Do these four things before session 1 on Saturday 14 November:

1. **A laptop** with a current web browser (Chrome works best). A phone is not enough for the labs. If you do not have a laptop, email the instructor before Thursday 12 November to ask about a loaner.
2. **A personal Google (Gmail) account.** The AI features in Google Colab need an account whose owner is 18 or older. Use a personal account, not an RSU or home-university account: school accounts often have Colab's AI switched off.
3. **The pre-course check.** The pre-course email (sent Monday 9 November) links to a short notebook called S0 and a 10-minute video. A notebook is a web page of small boxes of code, called cells, that you run one at a time; the video shows how. Open the notebook, run its first cell, ask Gemini (Google's AI assistant, built into Colab) the question written in it, and email a screenshot of Gemini's reply to the instructor by **Thursday 12 November**. If anything fails, say so in the email; it is much easier to fix before session 1 than during it.
4. **Install nothing.** Everything runs in your browser. You do not need a paid plan.

   Bring a pen to every session. Calculators are supplied in class.

Before session 3 (Thursday 19 November). The course now uses the tools marketers use on big data, all free, in the browser, on the same Gmail account:

1. **The course website, idb624.marxp.com.** One link, no account. It holds each session's slides and sheets, a practice warehouse where you ask the course data questions in SQL (the language a data warehouse is asked in; Gemini writes it from your steps), a checker for your hand cards, and the homework page.
2. **A BigQuery sandbox.** BigQuery is Google's data warehouse. Open it with your Gmail account and make a free sandbox project: no card, nothing to pay. Each pair needs at least one working sandbox in session 3.
3. **Before session 5 (Saturday 21 November):** open Data Studio (Google's free dashboard tool, called Looker Studio until 2026) once with the same account.

How this course works

The course runs over four weeks, on weekday evenings and Saturdays:

Session	Date	Time
1	Saturday 14 November	13:00-17:00 (lab)
2	Tuesday 17 November	18:00-21:00
3	Thursday 19 November	18:00-21:00
4	Saturday 21 November	09:00-12:00
5	Saturday 21 November	13:00-17:00 (lab)
6	Tuesday 24 November	18:00-21:00
7	Thursday 26 November	18:00-21:00
8	Saturday 28 November	09:00-12:00
9	Saturday 28 November	13:00-17:00 (lab)
10	Tuesday 1 December	18:00-21:00
11	Thursday 3 December	18:00-21:00
12	Tuesday 8 December	18:00-21:00
13	Saturday 12 December	09:00-12:00 (project clinic)
14	Saturday 12 December	13:00-16:00 (presentations and oral defences)

Every teaching session follows the same routine, so you always know what comes next:

1. **Problem.** A real business question, and the quick answer that goes wrong. You write down your own first guess.
2. **Hand.** You work the method by hand on your own hand card, with your laptop closed and your phone in your bag. Your neighbours have different cards.
3. **Ask.** You and your partner turn the steps you just did into a prompt for Gemini in Colab.
4. **Match.** You type your numbers and the AI's numbers into the notebook. It tells you whether they agree. Nothing goes to the full table until they do.

5. **Scale.** The AI runs your steps on the full data, and you run the checks: do the rows add up, do the totals agree, does one customer pulled back out of the result look right? Each session also hides one planted mistake for you to find.
6. **Tell.** You fill in a decision card: what you would do, two of your numbers, and what would change your mind.

After each session, work through its slides before the next session. The slides are written to be studied on their own, not just shown in the room: every page explains its idea in full, with every worked example step by step. A shorter companion chapter follows each session's slides. Each session also has an ungraded exercise sheet with answers, so you can check yourself.

Plan for about 12 hours of independent study each week beyond the scheduled sessions — the reading, the exercises, and the coursework live there.

Session by session

Session	Title	Topics
1	How much did we really sell? What big data is, and checking an AI assistant's number by hand	Why a spreadsheet cannot open our sales table, and what big data is: the three Vs of volume, velocity and variety, plus veracity, whether you can trust the number; Where a digital business gets its big data: the data sources a customer's trail leaves, and the four questions firms ask of them; A Colab notebook: running cells in order, and restarting when it breaks; Asking Gemini in plain steps, then inserting and running the code it writes; One row, one invoice line: what the sales table of a UK online gift seller counts; Revenue as quantity times price, worked on an eight-line hand card; The match: verifying an AI analysis against your hand card before trusting it on every invoice line; Course data only: what happens to every prompt you send
2	Two teams, two revenue totals: data cleaning with a ledger that reconciles	Rows that are not sales: cancellations, postage and fee lines; Blank customer IDs: what you can still say about sales without a customer; Duplicate lines and the overlap between the two yearly sheets; The ledger: rows in equal rows kept plus rows dropped, and revenue totals that reconcile; The mismatch protocol: what to do when your hand card and the AI disagree

Session	Title	Topics
3	“Sales collapsed in December 2011!”: asking a data warehouse for KPIs, and the partial-period trap	Big data needs a home: the data warehouse, one shared place for data that holds every invoice line and answers many people at once; A query: your steps, Gemini’s SQL, and a small answer table out of a million lines, first on the course website’s practice warehouse and then in BigQuery; Query cost: reading how much data a query will scan before it runs, on the course’s sales table and on a very large public table; A KPI as a numerator over a denominator: average order value is revenue per invoice, not per invoice line; The monthly trend and the partial-period trap: the data stops on 9 December 2011; A chart whose title states its message, checked against three hand-plotted points
4	Who gets the VIP voucher? Scoring every customer by recency, frequency and money (RFM), and letting an AI group them	From a million invoice lines to one row per customer: the data warehouse does the counting, and one customer pulled back checks it; Recency from a stated ’today’, frequency as invoices, monetary value net of cancellations; RFM scores from 1 to 5, and the named segments marketers act on; Clustering: an AI groups the customers that look alike, worked first for six customers by hand on a recency-by-money grid; Segments by rule against clusters by machine: who gets the voucher within the budget
5	Are we keeping the customers we win? Cohorts, retention, churn and a dashboard	A cohort: customers grouped by the month of their first purchase; The retention rate, worked as a ten-customer triangle by hand, then for every cohort in the data warehouse; Churn as an inactivity window you choose, and why it cannot run past the last date in the data; A dashboard: the small tables that sum up a million lines, kept open for a manager in Data Studio; Lab 2: the full-scale heat map and one action per RFM segment in the notebook, then the dashboard built from the sheet the notebook writes
6	Did the email work? A/B tests, control groups and uplift	“We emailed our best customers, so they spent more”: why that comparison proves nothing; The A/B test: a control group made by random assignment; Uplift: the gap between the emailed group and the control group; Noise and the shuffle test: could the gap be luck?; A success metric and a guardrail, chosen before looking at the results

Session	Title	Topics
7	What goes in “you may also like”? Basket analysis with support, confidence and lift	A basket: the products bought together on one invoice; Support and confidence for a product pair, worked on six baskets; Lift: is a pair bought together more often than chance would give?; Fake pairs: postage codes and best-sellers that look like everyone’s match; Basket rules versus recommender systems (overview)
8	Who visits but does not buy? Web sessions, conversion rates and the three Vs of marketing data	The three Vs from session 1, put to work: which V makes clicks, reviews and ad data hard to use; Who made the record: first-, second- and third-party sources a Thai online seller uses; Web analytics: a web session, and the conversion rate by traffic source and visitor type; What a conversion rate cannot show, and where to spend the next marketing budget; The final project brief: one marketing decision, analysed and defended
9	41,000 comments nobody can read: joining tables and AI labelling of customer reviews	A second company’s tables: orders, reviews and the key that links them; Joining tables without multiplying rows, checked with the ledger; LLM labelling of review text, checked against your own labels and the star ratings; LLM labelling at scale: sample, batch and cache, and why personal data never goes into a prompt; Lab 3: review themes joined to late deliveries
10	We can call 500 customers: which ones? Churn scores, base rates and customer lifetime value	A vendor’s ’97% accurate’ model, and the base rate that makes accuracy meaningless; Learning from the past and testing on a later period, so no future data leaks in; A churn score and the cut-off that picks 500 customers; Customer lifetime value at a declared margin; A confusion table priced at the call budget, against the simple recency rule
11	Can we send this list to the agency? Personal data, the PDPA, the GDPR and AI-tool terms	“We deleted the names, so it is anonymous”: re-identification from a list without names; Lawful basis for using customer data under Thailand’s PDPA; Profiling and decisions by score alone: the rights each law gives (GDPR Articles 21 and 22; the PDPA as checked against the course reference text); Who is accountable: the data controller and the data processor; AI-tool terms: what a marketing team may paste into an AI assistant

Session	Title	Topics
12	Which channel earned the sale? Funnels and credit	Event data and the funnel: from visit to purchase, step by step; Where visitors drop out of the funnel, read from the session-8 shop's Google Analytics sessions; Attribution: three channels claim the same sale, and credit is not cause; Last-click and data-driven attribution in GA4, with media-mix modelling named as the next step; Project check-in: one number, its hand check, and the test that would settle it
13	Project clinic: rehearsing the presentation and the oral defence	Rehearsing the presentation: one marketing decision, three checked numbers and the test that would settle it; Rehearsing the oral defence: redoing one step of your own analysis aloud with a changed number
14	The project: team presentations and oral defences	Pair and trio presentations: one marketing decision, its checked numbers and the test that would settle it; Individual oral defence: one step of your own analysis redone aloud with a changed number

Assessment

Component	Weight	Detail
Coursework	100%	4 graded pieces, set in sessions 2, 6, 8 and 11; the due dates follow below

Deadlines. 20:00 Bangkok time on the date each brief states: CW1 Sunday 22 November 2026, CW2 Sunday 29 November 2026, the project proposal Tuesday 1 December 2026 at the start of session 10, the project notebook and memo Friday 11 December 2026. Hand cards and the privacy audit are done and collected in class; presentations and oral defences are in session 14 on Saturday 12 December 2026.

Late penalty. 10% of the available marks per calendar day, zero after 3 days. The project notebook and memo take no late days, because the presentations are the next day.

Resubmission. None.

There is no exam. Your grade comes from four pieces of coursework:

Piece	Set in	Who does it	AI allowed?	Weight
CW1 Customer Value Pack: your best 3 of 4 hand cards from sessions 2-5, plus the Lab 2 product (customer segments, a retention table, one action per segment, a decision card)	Session 2	Pairs; each hand card is your own	Yes, except on hand cards	20%
CW2 Campaign and Review Insight Pack: your best 3 of 4 hand cards from sessions 6-9, plus the Lab 3 product (the email test read, the review labels checked by hand, a recommendation)	Session 6	Pairs; each hand card is your own	Yes, except on hand cards	20%
Final project: one marketing decision, chosen from five variants on the course's data sets. Proposal 5%, notebook and two-page memo 20%, presentation 10%, individual oral defence 15%	Session 8	Pairs or trios; the oral is yours alone	Yes, except in the oral	50%
CW3 Privacy audit: three fixes for a real campaign, tied to Thailand's personal data law	Session 11	Individual, in class, on paper	No	10%

Your **hand cards** are collected at the end of each hand-work block. They show your own thinking, before any AI was involved, so they are marked as your individual work even inside a pair's pack.

In the **oral defence** (3 minutes, in session 14) you sit alone and take four steps you know in advance: your memo's headline number with its unit and its count; how it was checked by hand; one step of your own analysis redone aloud with one number the examiner changes (you do not know which number); and the verdict, whether the call changes and at what number it would. You rehearse it in session 13. You may use your own glossary card, your oral card from session 13 and a basic calculator, not a phone; that is the course's default, and the instructor may change it. The memo, phones and AI tools are not allowed. The oral is marked on content, not on your English, and the examiner will rephrase any question on request. If you cannot explain your own headline number, the project mark is capped at half for you.

The **exercises** in every session are ungraded practice with worked answers. They are the fastest way to find out what you have not understood yet, before a graded piece finds it for you.

How to succeed

- **Do your hand card honestly.** It is the part of the course that teaches you the most, and a wrong hand card that you then correct is worth more than a right number you copied.
- **When your number and the AI's disagree, do not reword the prompt and hope.** Follow the mismatch protocol from session 2: recheck your arithmetic, ask the AI to show its working on your card's rows, and fix the step that differs. A real AI mistake that you catch and explain earns full marks for that part.
- **Work through each session's slides within two days of it.** The course moves fast: some weeks have four sessions, and each one builds on the last.
- **Keep a glossary card** of the terms in your own words, in English and, if it helps, in Thai. You may use it in the oral defence.
- **Ask early.** A question asked in session 3 saves a lost mark in session 10.

Materials

- **The session slides**, one deck per session: the main study text, written to be read on your own, with every worked example in full. A companion chapter, the exercise sheet and the Colab notebooks come with each session.
- **The data.** All public, anonymised data sets:
  - UCI Online Retail II: two years of sales from a UK online gift seller (CC BY 4.0). The course's main company.
  - The Hillstrom email test: a real email experiment with 64,000 customers (session 6).
  - UCI Online Shoppers Purchasing Intention: 12,330 website visits (session 8, CC BY 4.0).
  - Olist: orders and customer reviews from a Brazilian online marketplace (session 9). Reviews are shown with English translations.
  - Google's GA4 sample: website event data, queried in BigQuery later in the course.
  - Google Trends top search terms: a very large public table in BigQuery, about 200 million rows (session 3), to see what a big question scans.
- **The course website and the course's BigQuery tables.** The gift seller's lines are in one shared, read-only BigQuery dataset, and the same tables are on the course website.
- **Your laptop** and **your personal Google account** (see *Before you start*). Calculators are supplied in class.
- There is no textbook to buy.

Academic integrity

The rule for AI is the same in every brief:

> You may use Gemini in Colab in labs and coursework, but every number you submit must match your own hand card and a cell that ran. Hand cards, the privacy audit and the oral defence are done without AI, with phones in bags. Paste nothing personal into any AI tool.

- **Which AI.** Gemini in Colab is the course assistant. If it fails for you in a session, use your partner's account first, then the Gemini web app in a side window, then work on paper with your partner.
- **What you paste.** Use course data only. Google may keep what you type into Colab's AI, with the code and the output, for up to 18 months, and human reviewers may read it.
- **Pairs.** You and your partner share the work and the marks for the pair's product. Every pair submission ends with one line saying who did what. Copying another pair's numbers, or another student's hand card, is misconduct.
- **Your own words.** The memo and the decision cards are yours. An AI may help you check your English, but every claim must rest on a number you checked.

Contact

Watthanasak Jeamwatthanachai

Talk to the instructor before or after any session. Between sessions, ask on the course hand-in site, rsu.marxp.com. The course website is idb624.marxp.com.