

Kursbeschreibung: Data Literacy – Ausgewählte Kapitel 2023

(Stand: 08.03.2023)

1 Terminplan

KW	Datum	Uhrzeit	Raum	Dozent
15	Freitag, 14.04.2023	09:00 – 16:30 Uhr	WiWi-Pool 3	Dr. Walter Schmitting
16	Freitag, 28.04.2023	09:00 – 16:30 Uhr	WiWi-Pool 3	Dr. Walter Schmitting
17	Freitag, 05.05.2023	09:00 – 16:30 Uhr	WiWi-Pool 3	Simon Schölzel, M.Sc.
18	Freitag, 12.05.2023	09:00 – 16:30 Uhr	WiWi-Pool 3	Simon Schölzel, M.Sc.
19	Freitag, 19.05.2023	09:00 – 16:30 Uhr	WiWi-Pool 3	Dr. Walter Schmitting

2 Credits

3 CP / ECTS (2 SWS), Modul „Schlüsselqualifikation“

3 Dozenten

Dr. Walter Schmitting (Walter.Schmitting@wiwi.uni-muenster.de)

Simon Schölzel, M.Sc. (Simon.Schoelzel@wiwi.uni-muenster.de)

4 Ziel der Veranstaltung und Vorkenntnisse

Das Ziel dieser Schlüsselqualifikation ist es, Studierenden frühzeitig die wesentlichen Kompetenzen im Umgang und der Arbeit mit Daten („Data Literacy“) zu vermitteln. Insbesondere erlernen die Studierenden dabei strukturiert über die Arbeit mit Daten nachzudenken, Daten zu modellieren und zu interpretieren sowie die entsprechenden praktischen Tools und Technologien entlang des Data Science-Prozesses einzusetzen. Die Inhalte des Kurses bereiten Studierende auf eine potenzielle empirische Abschlussarbeit in der Wissenschaft sowie die Realität der Arbeit mit Daten in der Praxis vor. Theoretische Vortragseinheiten (video-on-demand) wechseln sich dabei mit anwendungsbezogenen Einheiten und Übungen zur Vertiefung des Gelernten ab. Studierende sammeln (erste) Erfahrungen im Umgang mit Excel, Python sowie Tableau. Der Fokus liegt hierbei auf einer überblicksartigen Wissensvermittlung und hat nicht den Anspruch eines zum Beispiel tiefgehenden Programmierkurses. Es wird erwartet, dass die Studierenden ein intrinsisches Interesse an den behandelten Themen mitbringen und sich aktiv einbringen. Die individuelle Lernkurve hängt dabei von der eigenen Motivation sowie den Vorkenntnissen der Studierenden ab und kann eine konsequente Nachbereitung der Inhalte zwischen der Veranstaltungsterminen erfordern. Generell sollten Teilnehmer:innen folgende Vorkenntnisse mitbringen:

1. Erfolgreicher Abschluss der Module „Statistik 1“ und „Statistik 2“, „Mathematik für Wirtschaftswissenschaftlicher“ und insbesondere „Technikern der IT“.
2. Vorkenntnisse im Umgang mit Python oder einer anderen Programmiersprache zur Datenanalyse sind nicht erforderlich. Gleiches gilt für Tableau.

Die Teilnehmerzahl ist beschränkt. Eine Bewerbung erfolgt über die [Bewerbungsmaske](#) der IVV Wirtschaftswissenschaften (bis zum 31. März 2023). Die Teilnahme an der Veranstaltung wird spätestens eine Woche vor Veranstaltungsbeginn bestätigt. Aktuelle Veranstaltungsinformationen finden sich zudem auf der Website des [Forschungsteam Berens](#).

5 Inhalte der Veranstaltung

Themen	Beschreibung
Einführung „Data Literacy“	Die Veranstaltung startet mit einer ausführlichen Beschreibung und Reflektion des Konzeptes der „Data Literacy“. Dabei wird zunächst der Begriff der „Daten“ geklärt. Anschließend wird die Motivation der Data Literacy untersucht. Sie wird definiert, gegen verwandte Konzepte abgegrenzt und mögliche Inhalte werden anhand des Kompetenzrahmens von Ridsdale et al. (2015) aufgezeigt. Fundierende oder verbundene Themen werden dabei im Zusammenhang erörtert. Letztendlich sollte jeder Teilnehmer den Begriff für sein Verständnis definieren können. Im zweiten Block wird dann der Frage nachgegangen, wo und wie sich über das Internet Informationen und Daten zu ausgewählten Themen finden lassen. Dabei gilt die Aufmerksamkeit sowohl der allgemeinen Recherche wie auch der Identifikation von Datenquellen. Abschließend wird im dritten Block exemplarisch einem besonderen Phänomen in der Datenanalyse nachgegangen.
Simulationsrechnung	Einführung in die Simulation. Anschließend Vorstellung eines Anwendungsfalles (Unternehmensplanspiel). Im Weiteren wird dann unter Verwendung von Excel und unter Nutzung der integrierten Programmiersprache Visual Basic for Applications (VBA) ein Simulationsmodell entwickelt. Die Ergebnisse werden analysiert und interpretiert. Zielführende Entscheidungsempfehlungen werden abgeleitet.
Datenvisualisierung	Im Rahmen einer Einführung wird die Gestaltung von Standard-Businessgrafiken (auch praktisch unter Nutzung von Excel) vermittelt. Es wird darüber hinaus auf „exotische“ Grafikformen sowie den Missbrauch grafischer Darstellungen eingegangen. Das Thema Visual Analytics bildet den Übergang zu einer Einführung in die Software „Tableau“, die nicht nur die Visualisierung, sondern auch die Datenanalyse unterstützt. Übungen zu den einzelnen Themen runden den Einblick in die Datenvisualisierung ab.
Datenmanipulation & Datenanalyse mit Python	<p>Die Veranstaltung beginnt mit einer Einordnung des Begriffs "Data Science" in den übergeordneten Kontext der "Data Literacy". Es folgt eine Diskussion des Anforderungsprofils des Data Scientists sowie ein Überblick über den generischen Data Science Workflow. Dabei wird aufgezeigt die Kenntnisse einer Programmiersprache (z.B. Python) zur Analyse von Daten heutzutage unabdingbar ist.</p> <p>Der zweite Teil der Veranstaltung beginnt mit einer grundlegenden Einführung in die Programmiersprache Python, gepaart mit diversen praktischen Beispielen und Übungen. Im Fokus steht die Arbeit mit Jupyter Notebooks zur Vertiefung der Inhalte über Datentypen, Datenstrukturen, Vergleichsoperatoren und Funktionen.</p> <p>Im dritten (Haupt-)Teil der Veranstaltung folgt zunächst eine kurze Diskussion verschiedener Datenmodalitäten (Tabellarische Daten, Text, Bild, Audio) bevor die Teilnehmer:innen die grundlegenden Funktionen zur Manipulation und Analyse tabellarischer Daten in Python kennenlernen. Themenschwerpunkte sind dabei der Datenimport, die primären Datenmanipulationen auf Zeilen- und Spalten-Ebene, die Berechnung und Interpretation von Statistiken sowie das Gruppieren und Mergen von Datensätzen. Analog zu Teil zwei wechseln sich dabei theoretische Einheiten mit hands-on Einheiten und Übungen ab. Alle Beispiele werden anhand eines realen Datensatzes und unter Verwendung der Python Library `pandas` verdeutlicht.</p>
Verarbeitung von Textdaten	Einführung in die maschinelle Verarbeitung von Textdaten, insb. Datenaufbereitung mittel Regular Expression (Regex) und quantitative Analyse von Textinhalten. Dabei werden beispielhaft einige wesentliche Methoden und Konzepte aus dem Bereich Natural Language Processing (NLP) vorgestellt.
Korrelation & Kausalität	Der Themenblock zielt auf die Verinnerlichung von Denkmustern ab, die es erlauben, Korrelationen von kausalen Zusammenhängen zu unterscheiden. Es werden einige Strategien zur Identifikation von kausalen Zusammenhängen besprochen, die es ermöglichen kausale Ursachen-Wirkungs-Beziehungen messbar zu machen.
Datenbeschaffung mittels API	Die letzte Vorlesungseinheit gibt einen kurzen Überblick über Programmierschnittstellen (sog. Application Programming Interfaces oder „APIs“), wo diese überall im Internet zu finden sind und wie sie zur Datenbeschaffung genutzt werden können. Es werden einige frei zugängliche Schnittstellen exemplarisch vorgestellt und angewandt (z.B. zum Zwecke der maschinellen Übersetzung, Transkription von Audio-Dateien oder Generierung von Musik-Empfehlungen).

6 Erworbene Kompetenzen

Neben dem übergreifenden Ziel der Vermittlung von Kompetenzen im Umgang mit Daten („Data Literacy“) sind die Teilnehmer:innen nach erfolgreichem Abschluss der Veranstaltung in der Lage

- die Relevanz von „Data Literacy“ im 21. Jahrhundert einzuordnen.
- strukturiert über Fragestellungen zu reflektieren, zu deren Beantwortung die Analyse von Daten entscheidend ist. Dies beinhaltet die eigenständige Datenbeschaffung, Datenbereinigung und -manipulation, Datenanalyse sowie Datenvisualisierung.
- mit modernen und praxisrelevanten Datenanalyse-Tools umzugehen, insbesondere Excel, Python und Tableau.
- zwischen assoziativen (Korrelationen) und kausalen Ursachen-Wirkungs-Beziehungen (Kausalität) zu unterscheiden und die Gültigkeit augenscheinlich kausaler Aussagen kritisch zu hinterfragen.
- die Wichtigkeit und den Wert von APIs (Programmierschnittstellen) im Internet zu erkennen und diese selbstständig zu verwenden.
- datengestützte Analysen zu verstehen und eigenständige empirische (Abschluss-)Arbeiten durchzuführen.

7 Prüfungsleistung

Für eine erfolgreiche Teilnahme und Anrechnung ist die Anwesenheit und aktive Mitarbeit an allen fünf Terminen sowie das Bestehen einer 60-minütigen Abschlussklausur erforderlich. Die Abschlussklausur wird voraussichtlich am Freitag, den 16. Juni 2023 stattfinden.