

# › EINFÜHRUNG

	Zeit und Ort	Kontakt
<b>Vorlesung</b>	Siehe Stundenplan	<b>Till Florian Kauffeldt</b>  > Email: <a href="mailto:florian.Kauffeldt@hs-heilbronn.de">florian.Kauffeldt@hs-heilbronn.de</a>  Sprechstunde > Individuelle Vereinbarung per Email

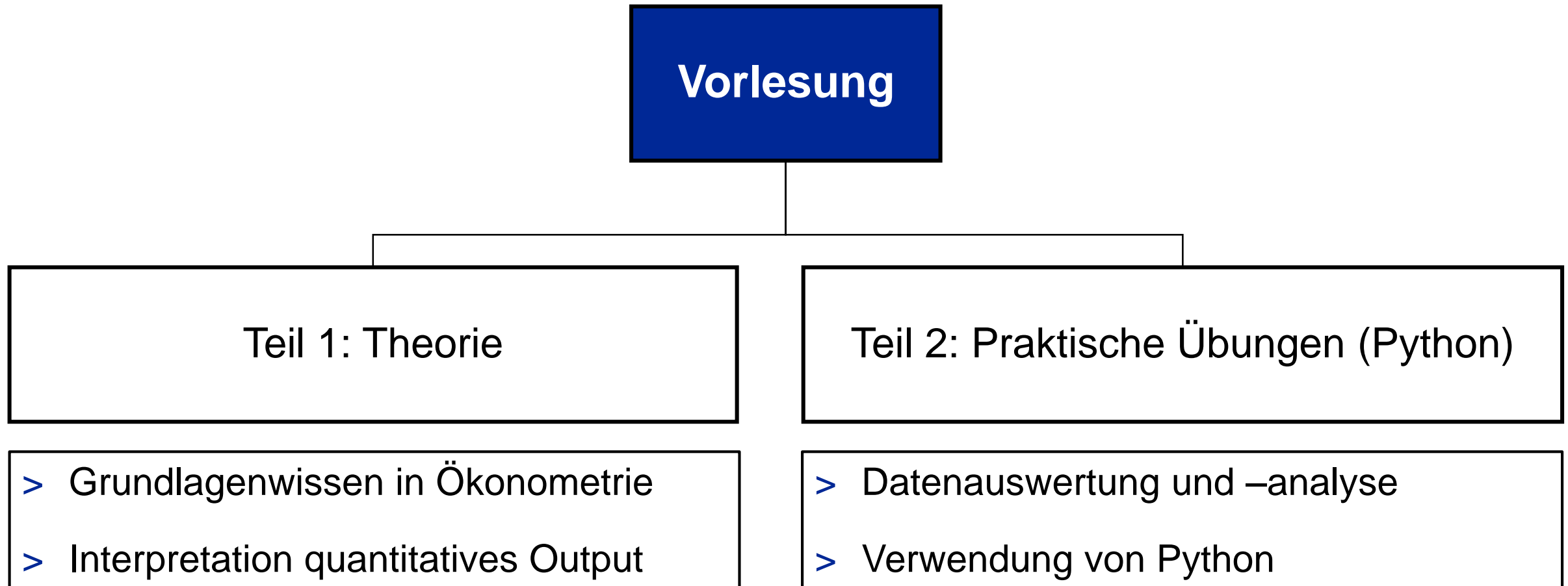
## 1. Einführung

## 2. Methoden

- > Deskriptive Methoden
- > Testmethoden für Parameter
- > Testmethoden für Korrelationen und Zusammenhänge
- > Multiples Testen

## 3. Modelle

- > Regressionsmodelle
- > KI-Modelle



# VORLESUNG: INTENDED LEARNING OUTCOMES (ILOS)

---

1. Possess essential knowledge in major fields of business to operate efficiently in a professional environment
2. Implement conceptual knowledge in business situations
3. Apply structured thinking and problem solving ability
4. Possess digital literacy and apply it to business environment
5. Work in teams and leverage diversity
6. Manage projects in a timely and efficient manner
7. Identify and critically assess business opportunities
8. Behave in an ethical and sustainable manner

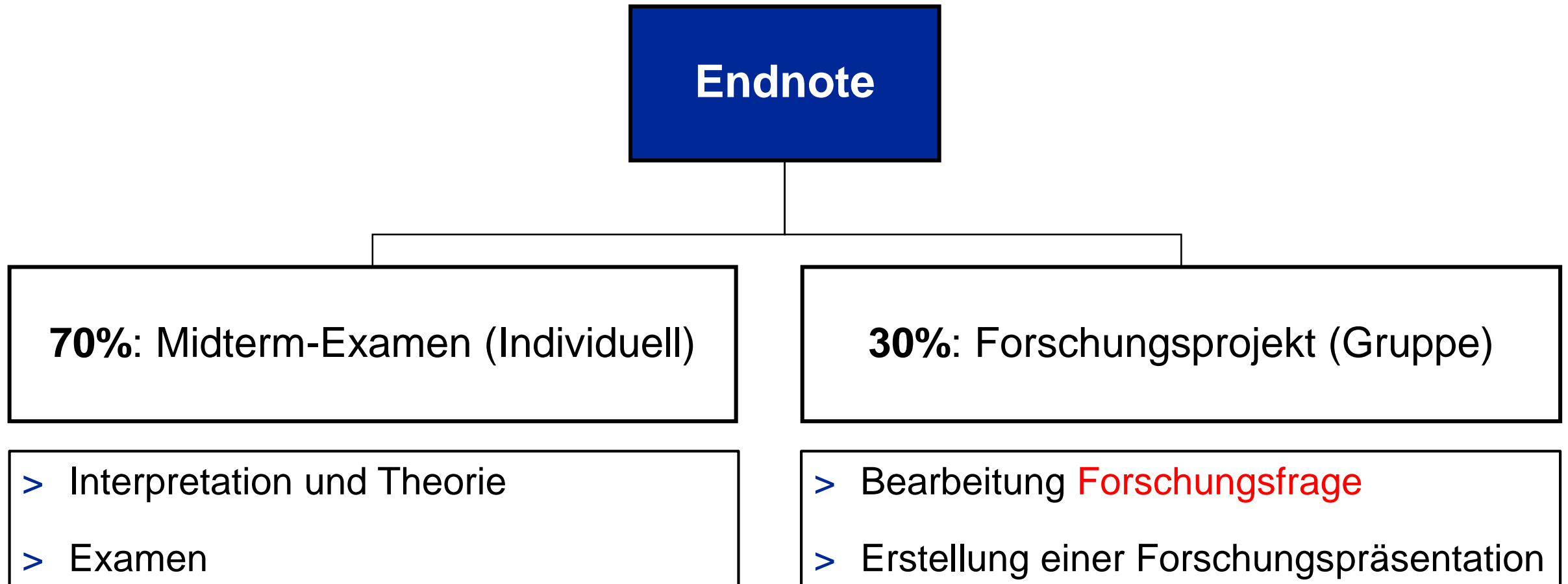
# VORLESUNG: VORKENNTNISSE

---

- > Mathematik (deutsches Abiturniveau)
- > Statistisches Grundlagenwissen (Statistik 1, 1. Semester)

Besonders wichtige statistische Grundlagen:

- > Messniveau (Nominal, Ordinal, Quantitativ/Nummerisch)
- > Quantile
- > Varianz / Standardabweichung
- > Null- und Alternativhypothese
- > Einseitige vs. zweiseitige und Einstichproben- vs. Zweistichproben-Tests
- > P-Wert
- > Power eines Tests





## 1. Entwicklung Forschungsfrage zu Forschungsdatensätze

> Werden noch in ILIAS bereitgestellt

## 2. Erarbeitung von geeigneten Hypothesen zur Beantwortung der Forschungsfrage

- > 2 bis 4 Hypothesen
- > Hypothesen dürfen nicht einfach aus dem "Nichts" kommen
- > Herleitung der Hypothesen: Existierenden Theorien (z.B. Psychologie, Wirtschaft,...)
- > Literaturquellen müssen angegeben werden → **Nur Peer-Reviewed Journals!**

## 3. Empirische Überprüfung der Hypothesen durch quantitative Methoden mit Python

- > Python-Codes müssen angegeben werden [.html oder .pdf]
- > Geeignete Quantitative Methoden: Voraussetzungen und Skalenniveau beachten!
- > Null- und Alternativhypothesen kontextbezogen bei Tests angeben!

## 4. Zusammenfassung der Schritte 1. bis 3. in einer Forschungspräsentation:

### Teil 1. Einführung

- > Hypothesen angeben und theoretischer Hintergrund nachvollziehbar erläutern
- > Literaturquellen (peer-reviewed) bei den jeweiligen Hypothesen aufführen

### Teil 2. Methodologie

- > Empirische Strategie (Verwendung der Quantitative Methoden) genau erläutern
- > Datensatz beschreiben

## 4. Zusammenfassung der Schritte 1. bis 3. in einer Forschungspräsentation:

### Teil 3. Resultate

- > Resultate (Output-Tabellen, Grafiken) aufführen und beschreiben
- > Zu jedem Test Null- und Alternativhypothese angeben (kontextbezogen)
- > Resultate interpretieren

### Teil 4. Schlussteil

- > Resultate zusammenfassen
- > Mögliche Probleme hinsichtlich der empirischen Strategie oder Datensatz beschreiben

## 4. Zusammenfassung der Schritte 1. bis 3. in einer Forschungspräsentation:

### Teil 5. Literaturverzeichnis

- > Wissenschaftliches Literaturverzeichnis → Auf die Form achten
- > Orientieren Sie sich an den Papers, die Sie lesen

### Teil 6. Appendix

- > Zusatzmaterialien (Grafiken, Zusatzinformationen,...) aufführen

## Zusammenfassung

1. Entwicklung Forschungsfrage zu einem der genannten Datensätze
  2. Erarbeitung von 2 - 3 geeigneten Hypothesen zur Beantwortung der Forschungsfrage
  3. Empirische Überprüfung der Hypothesen durch quantitative Methoden mit Python
  4. Zusammenfassung der Schritte 1. bis 3. in einer Forschungspräsentation
- > Einführung, Methodologie, Resultate, Schlussbemerkung, Literaturverzeichnis, Appendix

# FORSCHUNGSPROJEKT: HINWEISE ZU ANFRAGEN

---

1. Sie haben in Statistik 1 bereits gelernt wie man ein Forschungsprojekt durchführt
2. Hier wurden besprochen was eine Forschungsfrage und Hypothesen sind
3. Falls Sie dies nicht mehr wissen, müssen Sie das eigenständig nacharbeiten
4. Fragen dazu wie z.B. "Ist unsere Forschungsfrage gut?", "Wie erstellen wir die Hypothesen?",... werden nicht beantwortet!
5. Forschungsprojekt eigenverantwortlich durchführen!
6. Gegebene Freiheitsgrade nutzen!



# FORSCHUNGSPROJEKT: BEWERTUNGSKRITERIEN

Statistik	Vortrag und Präsentationsfolien	Literatur und Hypothesen	Mitarbeit
50%	20%	20%	10%
<ul style="list-style-type: none"> <li>&gt; Empirische Strategie?</li> <li>&gt; Methoden geeignet?</li> <li>&gt; Zweckmäßig?</li> <li>&gt; Annahmen überprüft?</li> <li>&gt; Python-Coding fehlerfrei?</li> </ul>	<ul style="list-style-type: none"> <li>&gt; Darstellung Methoden Resultate?</li> <li>&gt; Nachvollziehbar?</li> <li>&gt; Form des Vortrags? Einheitlichkeit, Graphiken nicht verschwommen...</li> </ul>	<ul style="list-style-type: none"> <li>&gt; Wissenschaftliche Quellen?</li> <li>&gt; Form des Literaturverzeichnisses?</li> <li>&gt; Hypothesen hinreichend begründet?</li> <li>&gt; Fehlerfrei?</li> </ul>	<ul style="list-style-type: none"> <li>&gt; Time-Keeping</li> <li>&gt; Wortmeldungen bei anderen Vorträgen</li> </ul>

## Hinweise:

1. Bei stark unzureichender Erfüllung einzelner Kriterien wird eine 5,0 vergeben

> **Unabhängig vom Erfüllungsgrad der anderen Kriterien**

Beispiel. Hypothesen wurden nicht begründet und keine Literatur, dann 5,0 – auch bei perfekten statistischen Auswertungen

2. Einzelne Gruppenmitglieder arbeiten nicht mit

> Sie können diese Gruppenmitglieder per Email melden

> Vorher den Mitgliedern eine Frist setzen → dann abweichende Individualbewertung

**VERY IMPORTANT**

# FORSCHUNGSPROJEKT: ABGABEMÖGLICHKEITEN

- > Abgabe der Forschungspräsentation und Notebook ausschließlich über ILIAS möglich!
- > Keine Email Abgabe!
- > Abgabe-Deadline beachten!
- > Verspätete Abgabe: Teilleistung wird mit 5,0 gewertet – auch bei wenigen Minuten Verspätung (keine Ausnahmen – computerbasiert)
- > Abgaben in einem gängigen Format (PowerPoint, pdf), sonst: 5,0

**VERY IMPORTANT**

# FORSCHUNGSPROJEKT: WAS MUSS ABGEGEBEN WERDEN?

---

Bis **23.06.2024, 23:30**, muss folgendes abgegeben werden:

- > Forschungspräsentation (PowerPoint)
- > Python-Notebook mit Codes (Python-Notebook) **im html-Format**

- > Um ein Forschungsprojekt zu bearbeiten, bilden Sie selbständig Gruppen
- > Maximal 7 Teilnehmer → Alle Mitwirkenden auf der Präsentation angeben
- > Bei mehr als 7 Teilnehmern, werden nur die ersten 7 gewertet

Es ist Ihre Verantwortung sich einer Forschungsgruppe anzuschließen

- > falls Sie keine Gruppe finden, können Sie dies bis zum **10.04.2024** melden
- > falls Sie sich keiner Gruppe anschließen, müssen Sie das Projekt alleine bearbeiten oder erhalten eine 5,0 bei Nicht-Abgabe

# VORLESUNG: WICHTIGE TERMINE (OHNE GEWÄHR – ILIAS BEACHTEN)

- > 03.04.2024 (in der Vorlesung):  
**Deadline Verbindliche Anmeldung** (nicht angemeldete werden aus dem Kurs entfernt)
- > 10.04.2024: Letzte Möglichkeit Meldung, falls keine Forschungsgruppe gefunden wurde
- > 12.06.2024 (in der Vorlesung): Midterm-Examen
- > 23.06.2024 (23:30):  
**Deadline** Abgabe Forschungspräsentation+Python Notebook (bis 23:30)
- > 26.06.2024 (in der Vorlesung): Vorträge

**VERY IMPORTANT**

# ABGABE VERBINDLICHE ANMELDUNG

> Verbindliche Anmeldung:

> Formular wird in der Vorlesung ausgeteilt

> Deadline: 03.04.2024

> Ohne Unterschrift werden Sie aus dem Kurs entfernt

## Vorlage V

### Verbindliche Anmeldung / Prüfungsversuch

Binding registration/ examination attempt

Name der Lehrveranstaltung	Models and Methods of Data Assessment
Semester Semester	
Dozent/ Prüfer Lecturer	Prof. Dr. Florian Kauffeldt
Regeln/ Bedingungen der Lehrveranstaltung Lecture	Die Note setzt sich wie folgt zusammen: 50% Midterm-Examen (individuell) 50% Forschungsarbeit (kollektiv)

# VORLESUNG: AKADEMISCHE UND PRAKTISCHE RELEVANZ



The world is now awash  
in data and we can see consumers  
in a lot clearer ways.

- Max Levchin  
Co-founder of PayPal



Information is  
the oil of the 21st century,  
and analytics is  
the combustion engine."

- Peter Sondergaard  
( Gartner Research )



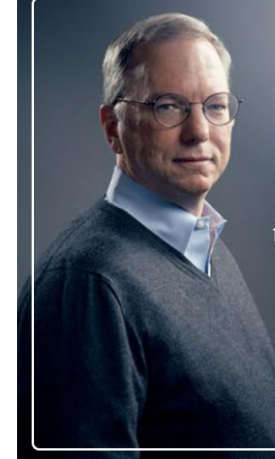
EVERY COMPANY HAS BIG DATA IN  
ITS FUTURE AND EVERY COMPANY WILL  
EVENTUALLY BE IN THE DATA BUSINESS.

Thomas H. Davenport  
President's Distinguished Professor in IT  
and Management



There were 5 exabytes of information  
created between the dawn of civilization  
through 2003, but that much information  
is now created every 2 days.

- Eric Schmidt  
Executive Chairman of Google





# VORLESUNG: AKADEMISCHE UND PRAKTISCHE RELEVANZ

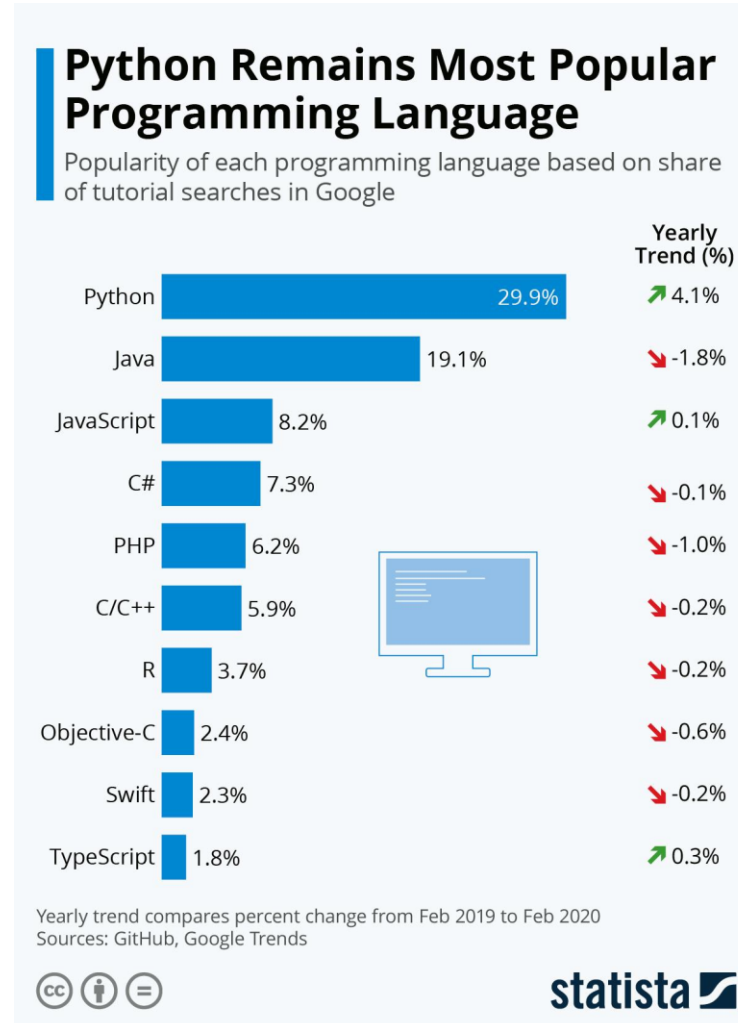
---

Grundlagenwissen Datenanalyse und Interpretation ist unabdingbar:

- > Statistik = Lernen aus Daten
- > Unternehmen: Bestmögliche Entscheidungen basierend auf vorliegenden Daten
- > Echtzeitverarbeitung großer Datenmenge durch Computerisierung möglich
- > Neue Technologien: Big Data Analytics, Machine Learning

# PYTHON: EINE GROÖE BEREICHERUNG

## Warum Python?



# PYTHON: EINE GROÖE BEREICHERUNG

Warum Python?



[Lidl Digital International GmbH & Co. KG](#)

## Werkstudent Omnichannel eFood (m/w/d)

Neckarsulm Studentenjobs, Werkstudent Teilzeit

Erschienen: vor 1 Woche

### Dein Profil

- Studium der BWL, idealerweise mit einer Spezialisierung in Wirtschaftswissenschaften (Strategie/Digital & E-Commerce) oder Wirtschaftsinformatik
- Erste praktische Erfahrung in den Bereichen Digital Marketing & Strategy, Customer Journey und Experience Management, Omni-Channel Strategy, E-Commerce - Food/Non-Food wünschenswert
- Kenntnisse in Python und SQL von Vorteil

# PYTHON: EINE GROßE BEREICHERUNG

## Warum Python?



### Analyst im Online Marketing (m/w/d)

Würzburg

ab sofort | Bayern | XXXLutz E-Commerce Würzburg | Vollzeit

### Ihre Qualifikationen

- ausgeprägte analytische und konzeptionelle Fähigkeiten sowie eine selbständige Arbeitsweise
- die Fähigkeit, Zahlen, Daten & Fakten zu verarbeiten und zielgruppengerecht zu erklären & zu präsentieren
- Erfahrungen im Online-Marketing mit Webanalyse sowie Vorkenntnisse in Marketingtracking & -Analyse von Vorteil
- idealerweise Erfahrungen mit Technologien mit AdServer und/oder Google Marketing Platform; Python oder R Kenntnisse sind von Vorteil

# PYTHON: EINE GROßE BEREICHERUNG

Warum Python?



Wir als Arbeitgeber

Schüler

Studenten

Berufseinsteiger

Berufserfahrene

Jobbörse

Mein Kandidatenprofil

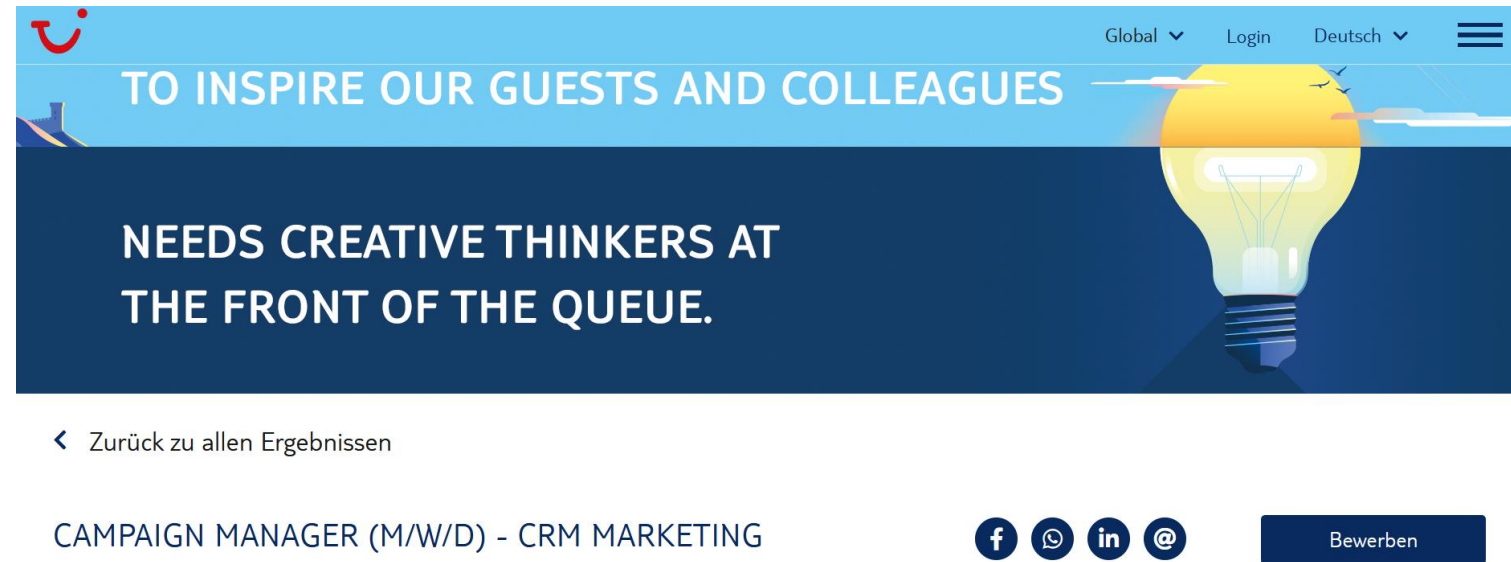
## Praktikant (m/w/d) Customer & Retail Analytics

Ort: 50933 Köln | Vertragsart: Vollzeit, befristet | 6 Monate | Job-ID: 240602

- Erste Kenntnisse im Einsatz von mindestens einem Datenanalyse-Tool (z.B. SQL, R, KNIME, Python) sind von Vorteil.
- Sie besitzen fundierte Fähigkeiten im Umgang mit Excel und Powerpoint.
- Eine selbstständige Arbeitsweise und eine schnelle Auffassungsgabe zeichnen Sie aus.

# PYTHON: EINE GROÖE BEREICHERUNG

## Warum Python?



### Wonach wir suchen

- Abgeschlossenes Studium in Wirtschaftswissenschaften, Marketing, Statistik, Wirtschaftsmathematik,-informatik oder eine vergleichbare Qualifikation, jeweils mit Schwerpunkt in quantitativen analytischen Verfahren
- Erste einschlägige Berufserfahrung im CRM / E-Commerce Umfeld im E-Mail-, Push- oder In-App-Marketing
- Verständnis für die CRM-Wertschöpfungskette und deren Zusammenhänge in technischer und wirtschaftlicher Hinsicht
- Erfahrungen im Umgang mit Kampagnenmanagement-Systemen (z.B. Braze, Salesforce, Adobe, IBM o.ä.)
- Erste Erfahrungen im Umgang mit Cloud-basierten Datenbank-Systemen (z.B. Snowflake, Redshift / AWS) oder klassischen Datenbanken (z.B. MS SQL, DB2) und entsprechenden Programmiersprachen (z.B. SQL, **Python**) oder vergleichbaren Technologien

# IHRE AUFGABEN BIS NÄCHSTE WOCHE

---

**Sehr wichtig: Installieren Sie Python!!!**

**VERY IMPORTANT**

- > Gehen Sie zu ILIAS → DOWNLOAD PYTHON-UNTERLAGEN
- > Nächste Woche erhalten Sie Unterstützung, falls Sie die Installation nicht hinbekommen
- > Danach müssen Sie dies selbständig lösen



# IHRE AUFGABEN BIS NÄCHSTE WOCHE

> Lesen Sie mindestens die ersten 3 Kapitel des Dokuments + Aufgaben lösen (3.3):

## Content

1	Installation .....	2
2	Hello World .....	5
2.1	Anaconda .....	5
2.2	VSC .....	6
3	First Steps with Data .....	7
3.1	Code Description .....	7
3.2	Loading Data .....	8
3.3	Exercises .....	8



# WIEDERHOLUNG: WICHTIGE STATISTISCHE GRUNDLAGEN

---

- > Messniveau (Nominal, Ordinal, Quantitativ/Nummerisch)
- > Quantile
- > Varianz / Standardabweichung
- > Null- und Alternativhypothese
- > Einseitige vs. zweiseitige und Einstichproben- vs. Zweistichproben-Tests
- > P-Wert
- > Power eines Tests

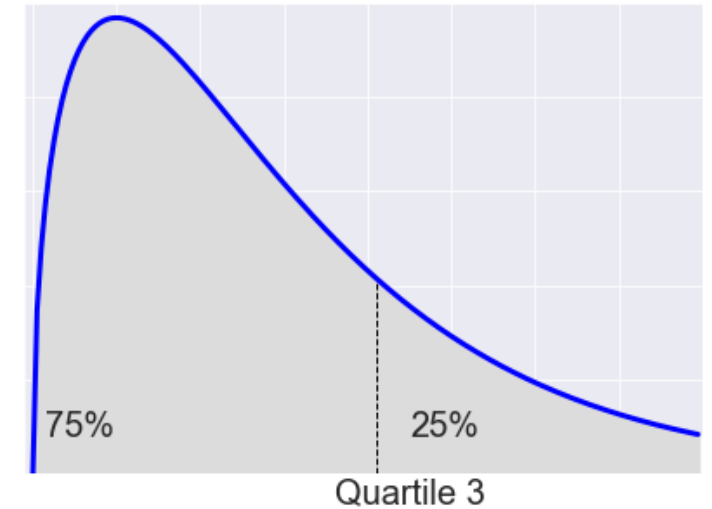
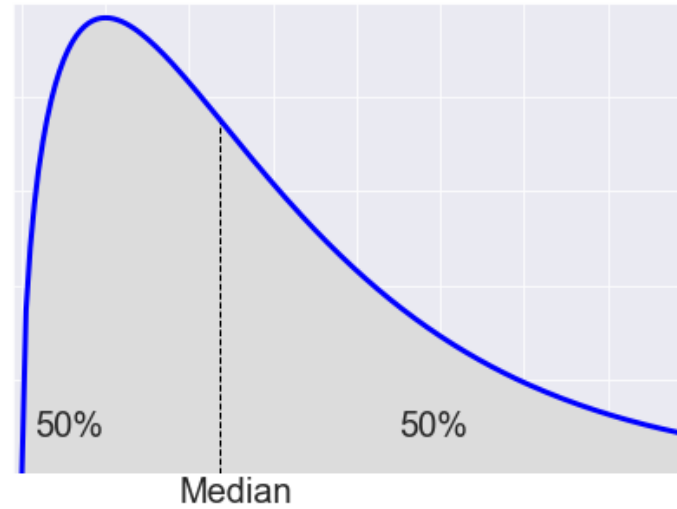
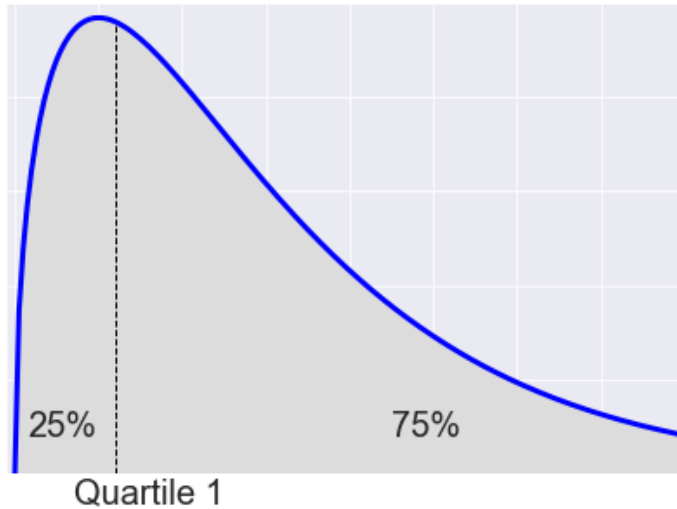
## **Nummerisches / Quantitatives Skalenniveau:**

- > Daten werden in Zahlen gemessen (Beispiele: Größe, Temperatur)
- > Mittelwerte können gebildet werden

## **Kategoriales / Qualitatives Skalenniveau**

- > Daten werden nicht in Zahlen gemessen (Beispiele: Militärränge, Farben)
- > Mittelwerte dürfen nicht gebildet werden
- > Nominal = Kategorien ohne natürliche Ordnung (Beispiel: Farben)
- > Ordinal = Kategorien mit natürlicher Ordnung (Beispiel: Militärränge)

# WIEDERHOLUNG: QUARTILE



# WIEDERHOLUNG: VARIANZ UND STANDARDABWEICHUNG

---

- > Varianz = Mittlere quadratische Abweichung vom Mittelwert im Datensatz
- > Standardabweichung = Mittlere Abweichung im Datensatz (positive Wurzel der Varianz)

Maße für die Streuung der Daten um Ihren Mittelwert

# WIEDERHOLUNG: EINSEITIG VS ZWEISEITIGER UND EIN- VS. ZWEISTICHPROBEN

- > *Einseitige Tests* haben eine Richtung: mehr/weniger, höher/niedriger, ...
- > *Zweiseitige Tests* haben keine Richtung: Unterschied ja/nein

**Beispiel.** Überprüfen Sie, ob die folgenden Tests einseitig oder zweiseitig sind:

Wir möchten wissen...	Alternativhypothese ( $H_a$ )	Nullhypothese ( $H_0$ )
ob Absolventen der HS Heilbronn im Schnitt ( $\mu_{Alter}$ ) älter sind als 26 ( $\mu_0$ )	$\mu_{Alter} > 26$	$\mu_{Alter} \leq 26$
ob sich der Durchschnitts-IQ ( $\mu_{IQ}$ ) von Studierenden verändert hat (langfristiger Durchschnitt: $\mu_0 = 101$ )	$\mu_{IQ} \neq 101$	$\mu_{IQ} = 101$
ob die Nachfrage ( $\mu_{Nach}$ ) für das Produkt einer Firma abgenommen hat (langfristiger Durchschnitt: $\mu_0 = 41'000$ )	$\mu_{Nach} < 41'000$	$\mu_{Nach} > 41'000$

## P-Wert

Der *p-Wert* eines Tests ist die bedingte Wahrscheinlichkeit (unter der Nullhypothese), dass die Teststatistik einen mindestens so extremen Wert annimmt wie denjenigen, der beobachtet wurde:

$$p - \text{Wert} = \Pr(\text{beobachtet oder extremer} \mid H_0 \text{ wahr}).$$

Wenn **p-Wert** < 5%, dann können wir die Nullhypothese ablehnen

> D.h. wir haben einen Beleg für die Alternativhypothese

## Power

Die **power** eines Tests ist die bedingte Wahrscheinlichkeit (unter einer spezifischen Alternativhypothese), dass der Test richtigerweise die Nullhypothese ablehnt:

$$power = \Pr(H_0 \text{ ablehnen} \mid H_a \text{ wahr}).$$

- > power = Wahrscheinlichkeit Erfassung wahr positiver Fälle
- >  $1 - \text{power}$  = Wahrscheinlichkeit Typ II Fehler (falsch-negativ)

gegeben, dass es einen Effekt gibt

# THE END!



Please refer any questions to:  
Prof. Dr. Florian Kauffeldt  
Faculty of International Business  
[florian.kauffeldt@hs-heilbronn.de](mailto:florian.kauffeldt@hs-heilbronn.de)