

# DIH SÜD - DIGITAL INNOVATION HUB SÜD

## *Deskriptive Datenanalyse Teil 2*

### *Anwendung und Umsetzung*

28.09.2022



Der DIH SÜD wird  
unterstützt von:



LAND  KÄRNTEN

# ***Deskriptive Datenanalyse***

---

## ***Anwendung und Umsetzung***

**FRANZ MOSER & HERMANN KATZ**

Institut für Wirtschafts- und Innovationsforschung POLICIES  
Forschungsgruppe „Datenanalyse und statistische Modellierung“

JOANNEUM RESEARCH Forschungsgesellschaft mbH

hermann.katz@joanneum.at

Graz, 28. September 2022

# Zeitplan

---

09:00 – 09:30: Vorstellung und Erwartungshaltungen

09:30 – 10:30: Vorgangsweise bei datengestützten Fragestellungen

10:30 – 11:00: Kaffeepause

11:00 – 12:00: Use Case 1

12:00 – 13:00: Mittagspause

13:00 – 14:00: Use Case 2

14:00 – 14:30: Kaffeepause

14:30 – 15:30: Use Case 3

15:30 – 16:00: Zusammenfassung und Resümee



# VORGANGSWEISE BEI DATENGESTÜTZTEN FRAGESTELLUNGEN

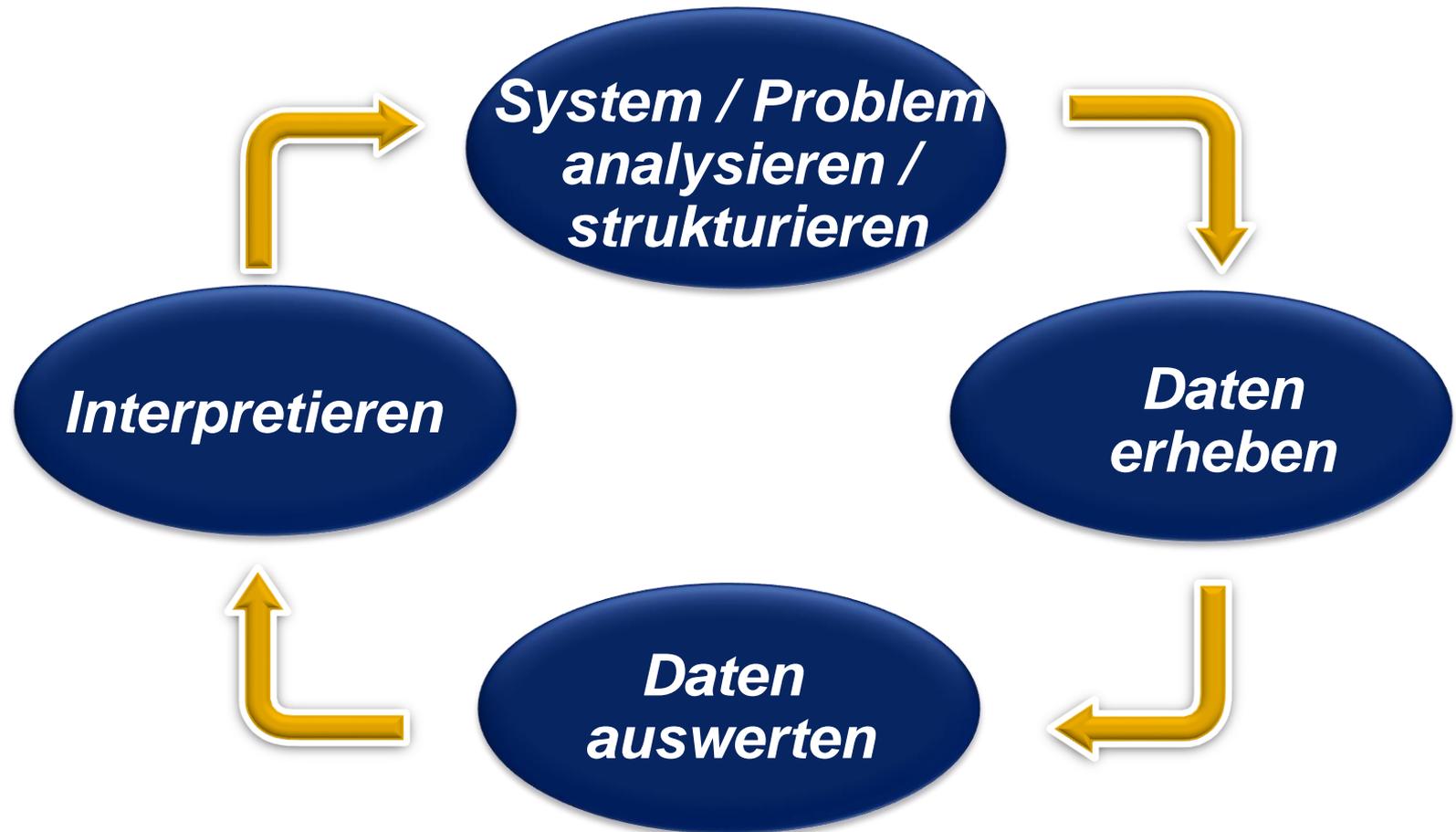


Der DIH SÜD wird  
unterstützt von:



LAND  KÄRNTEN

# Wissenschaftliche Vorgangsweise in der Angewandten Statistik



# Unser Vorgehensmodell ...



**Get**



**Explore**



**Model**



**Communicate**

- Was ist die spezifische Fragestellung?
- Analysieren der Anforderungen
- Analysieren des Prozesses bzw. des Systems

- Daten aus bestehenden Quellen gewinnen
- Neue Datenquellen erschließen
- Daten fusionieren und prozessieren

- Plausibilität prüfen
- Analysieren
- Visualisieren

- Detektion
- Klassifikation
- Prognose
- Optimierung
- Validierung

- Neue Erkenntnisse vermitteln
- Präsentation
- Report
- Software-Tool

# Systemanalyse - Planung

---

- Formulierung der sachspezifischen Fragestellung
  - Ziele festlegen
  - Fachwissen mit statistischem Know-how verbinden
- Merkmale festlegen
  - Skalierung, Eigenschaften
- Datenquellen konkretisieren
  - Fragebogendesign
  - Pretest
  - Projektdurchführung
- Grundgesamtheit – Stichprobe
  - Auswahl an einer repräsentativen Stichprobe



# *Datengewinnung*

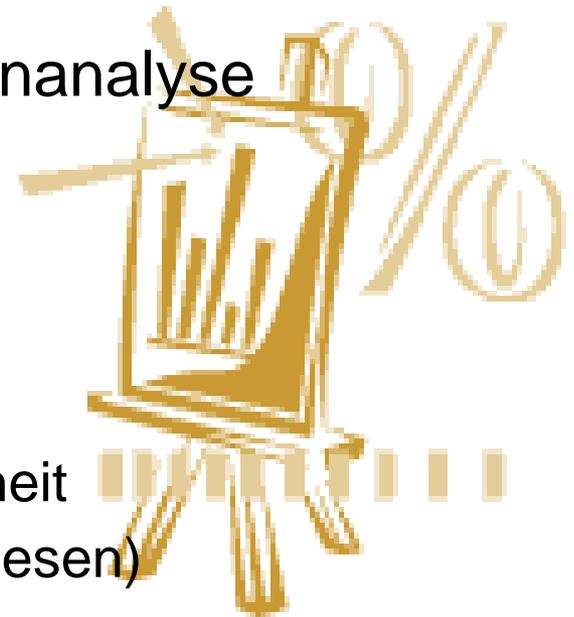
- Erhebungs- oder Versuchsplanung
  - Analyse der Auswahlgrundlage
  - Fragebogendesign
  - Pretest
  - Versuchsplanung
  - Festlegung des Stichprobenumfangs
  - Organisatorischer Ablauf der Datenerhebung – ev. Pilotstudie
- Datensammlung
- Dateneingabe – Online-Befragungen
- Überprüfung der Korrektheit der Daten



# *Statistische Auswertung*

---

- Kritische Analyse der Urdaten
- Deskriptive und exploratorische Datenanalyse
  - Tabellen
  - Grafiken
  - Kennzahlen
- Inferenzstatistische Aussagen
  - Aussagen bezüglich der Grundgesamtheit
  - Überprüfung von Vermutungen (Hypothesen)
  - Modellierung z. B. Regression



# *Sachspezifische Entscheidungsfindung*

- Aufbereitung der statistischen Ergebnisse für Entscheidungsfindung
- Sachspezifische Interpretation der Ergebnisse
- Ableitung von Maßnahmen
- Ev. Detailstudien



# KFV Standard Reporting



## Ask

Auswertung und Darstellung bemerkenswerter Aspekte der österr. Unfallstatistiken

## Get the Data

KFV, Statistik Austria, internationale Datenbanken

## Explore the Data

Tabellen und Grafiken

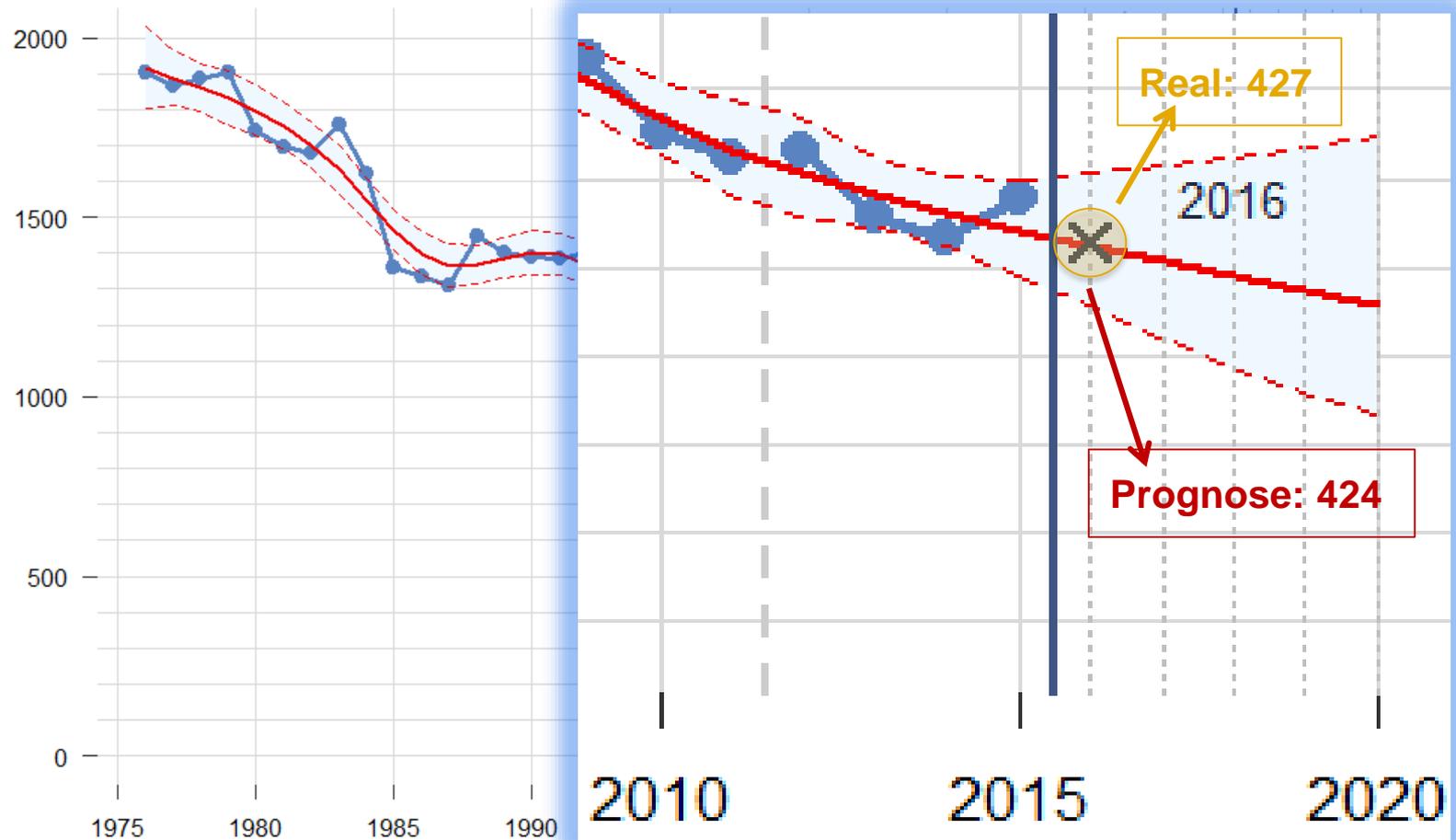
## Model the Data

Vorhersagen in Zukunft

Communicate and visualize the result

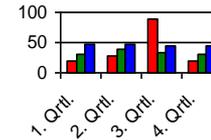
Bericht

# KFV Standard Reporting - Prognose der Todesfälle im Straßenverkehr

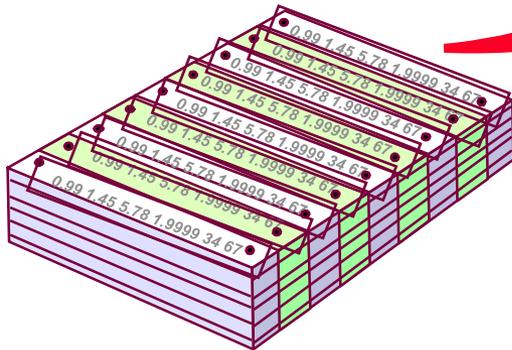


# Daten $\neq$ Information

**Information**



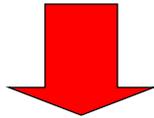
**Statistische Werkzeuge**



# ***Gliederung der Statistik***

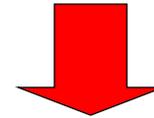
---

**Beschreiben**



**Deskriptive Statistik**

**Schlüsse ziehen**



**Inferenzstatistik**

# Deskriptive Statistik

## ■ Deskriptive (beschreibende) Statistik

- Instrumentarium zur Beschreibung von Daten
- Vorstufe zur schließenden Statistik

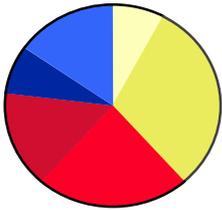


**Ziel:** Beschreibung, Strukturierung,  
Verdeutlichung, Darstellung  
umfangreichen, unübersichtlichen  
Datenmaterials

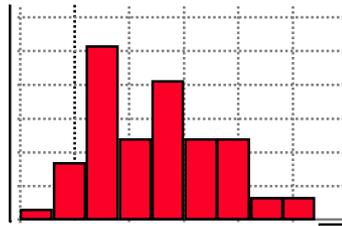
## ■ Methoden:

- Grafische Darstellungen
- Kennzahlen (Maßzahlen)

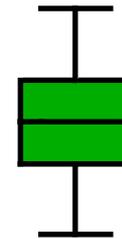
# Unterschiedliche Grafiken



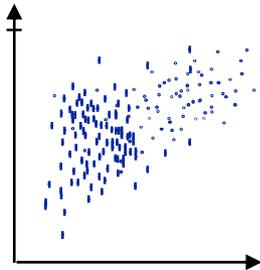
Piechart



Histogramm



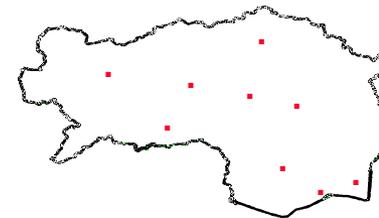
Boxplot



Scatterplot



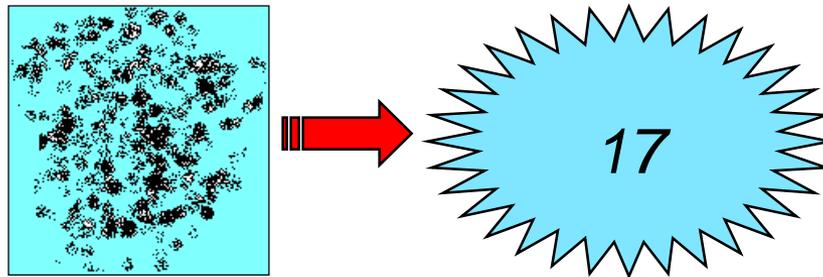
Zeitreihe



Map

# Allgemeine Kennzahlen

*Sinn von Kennzahlen ...*



- *Reduzierung von Komplexität*
- *Verdichtung von Information*

***... die Sache auf den Punkt bringen!***

# Skalierung

Skalentyp	Nominalskala
Definierte Relationen und Operationen	$\neq =$
Zulässige Transformationen	bijektive Transformation wie Umbenennen, Permutation
Beispiele für Merkmale	Familienstand, Geschlecht, Postleitzahl, Artikelbezeichnung, Religionszugehörigkeit
Merkmalsausprägung	Namen, Symbole, Codes



rot



blau



grün

# Skalierung

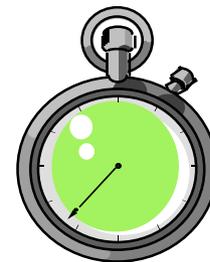
Skalentyp	Ordinalskala
Definierte Relationen und Operationen	$\neq = < >$
Zulässige Transformationen	isotone oder rangerhaltende Transformationen
Beispiele für Merkmale	Zeugnisnoten, Sozialstatus, Produktgüteklassen, Mercalli-Erdbebenskala
Merkmalsausprägung	Ordinalzahlen (i.d.R. ganze Zahlen)

gut  
besser  
G . . . . .

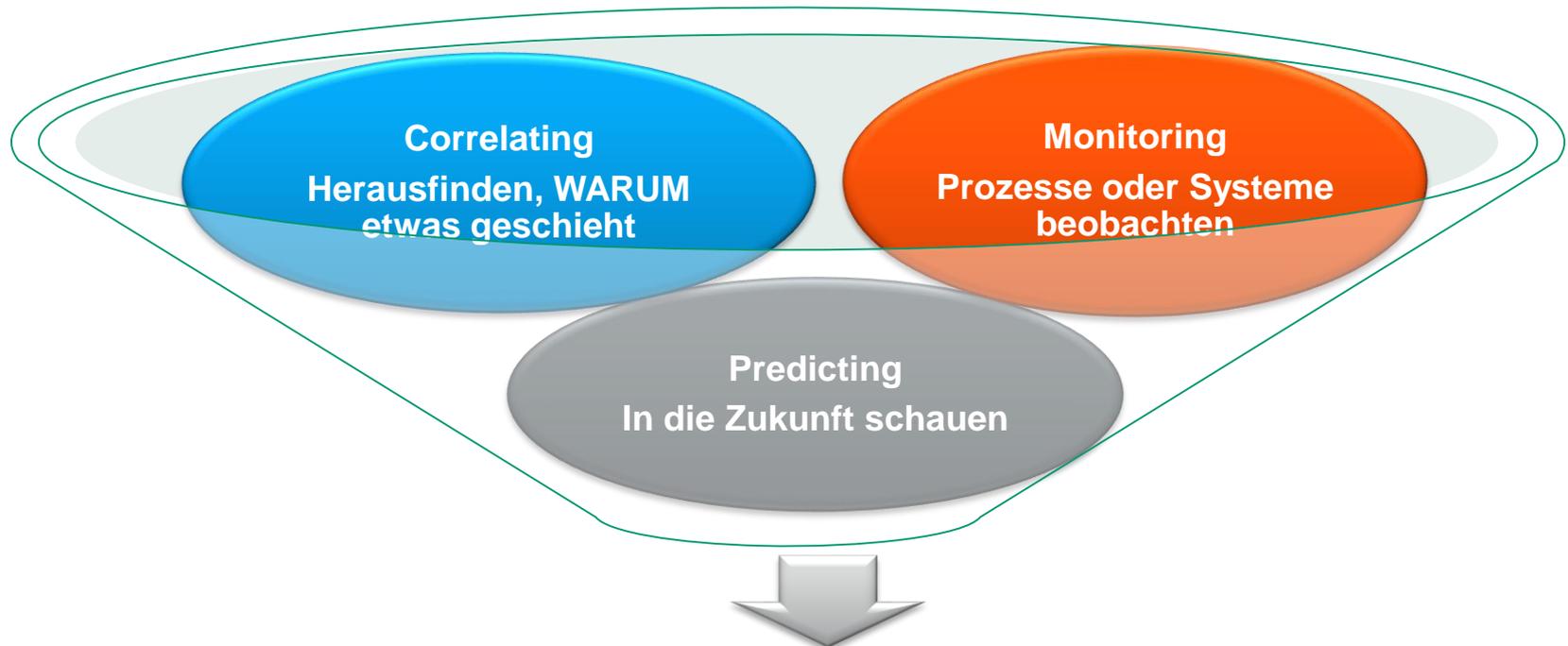


# Skalierung

Skalentyp	Kardinalskala (Metrische Skala)
Definierte Relationen und Operationen	$\neq = < > + - * /$
Zulässige Transformationen	Ähnlichkeitstransformationen $y=ax$ mit $a>0$
Beispiele für Merkmale	Temperatur, geographische Höhe, Messungen im cm-g-sec-System, Anzahlen
Merkmalsausprägung	reelle Zahlen (stetig oder diskret)



# Die Möglichkeiten statistischer Methoden ...



- Qualität und Zuverlässigkeit garantieren
- Durchsatz erhöhen
- Kosten reduzieren
- Verkaufszahlen und Gewinne steigern
- Bessere Entscheidungen treffen

# Präsentation Use Case 1



Der DIH SÜD wird  
unterstützt von:



LAND  KÄRNTEN

# Systemanalyse

---

- Insta Communications GmbH bietet psychologische Online-Beratung an
- Forschungsfrage: Was sind die wichtigsten Einflussfaktoren, die die Länge der Beratung (Anzahl an Beratungen) positiv beeinflussen?
- Die Identifikation von Faktoren, die sich fördernd oder hemmend auf die Beratungslänge auswirken
- Ansatz aus statistischer Sicht: Modellierung der vorhandenen Daten mit Zielgröße „Länge der Beratung“ (gemessen in Minuten oder Anzahl der Beratungsterminen)

# *Datensituation*

---

- Datensatz aller gestarteten Beratungen der letzten Monate
- Alter (als Faktor)
- Geschlecht
- Beratungsthema
- Qualität der Beratung durch die Psycholog:innen gemessen am Bewertungsscore
- Zeitpunkt der Registrierung bzw. des Beratungsstarts (Wochentag, Uhrzeit, Saison, Jahreszeit, Feiertage)
- gewähltes Beratungsmodell (täglich, wöchentlich, monatlich)
- Beratungsform (Video-, Audio- oder Textchat)
- Start als Selbstzahler oder Firmen- bzw. Versicherungskunde

# Präsentation Use Case 2



Der DIH SÜD wird  
unterstützt von:



LAND  KÄRNTEN

# *Systemanalyse / 1*

---

- Kreisler GmbH betreibt eine Elektrotankstelle mit angeschlossenem teilautomatisierten Lebensmittelladen.
- Der Zusammenhang zwischen Schnellladen und dem Ausfall des Stromspeichers soll analysiert werden.
- Der Speicher lädt bis 100%. Die einzige Stromquelle ist die PV-Anlage (kein Netzbezug).
- Wenn die Sonne scheint und der Ladezustand des Speichers nicht steigt (außer er hat 100% erreicht) gilt er als ausgefallen.

# *Systemanalyse / 2*

---

- Zielsetzung ist die Identifikation der Speicherausfälle bzw. der zugehörigen Zeiträume
- Analyse von möglichen Ursachen, die zu einem derartigen Speicherausfall führen
- Ansatz aus statistischer Sicht: Zusammenführung der drei Datensätze (Photovoltaikdaten Speicherdaten und Ladedaten) und Identifikation der Ausfallszeiten. Deskriptive Analyse der Daten, um mit Unterstützung von zugehörigen Grafiken die Ausfallszeiten zu erkennen.

# *Datensituation*

---

- Zeitraum der Daten bezieht sich auf die ersten vier Monate des Jahres 2021 und 2022
- Photovoltaikdaten
- Speicherdaten
- Ladedaten

# Präsentation Use Case 3



Der DIH SÜD wird  
unterstützt von:



LAND  KÄRNTEN

# *Systemanalyse / 1*

---

- Nuki Home Solutions GmbH steht für nachrüstbare Zutrittslösungen
- Nuki verschickt aktuell Waren in alle EU Länder und einige ausgewählte Länder im weiteren Ausland. Dieser Versand wird mit DHL (für deutsche Sendungen) und mit UPS (für alle anderen Sendungen) durchgeführt. Dabei kommt es immer öfter vor, dass Pakete nicht zugestellt werden können, weil der Kunde angeblich nicht zu hause war. Unser Supportteam bekommt dann die Meldung vom Kunden, dass er zuhause war, aber niemand gekommen ist.

# Systemanalyse / 2

---

- Wie ist die Verteilung dieser Problemfälle in den ausgewählten Ländern?
- Welche Länder/Städte sind hier besonders betroffen?
- Gibt es eine Auffälligkeit bei bestimmten Tagen (zB. Freitags, weil der Paketdienst früher ins Wochenende will)?
- Ist die Anzahl der Problemfälle im Verhältnis zu den Sendungen gestiegen, oder weicht das Verhältnis ab (in positive, oder auch negative Richtung)?
- Gibt es eine große Abweichung zwischen deutschen Städten (DHL) und anderen Städten (UPS)?
- Folgende Länder sollen analysiert werden: Deutschland, Österreich, Italien, Spanien und Frankreich
- Ansatz aus statistischer Sicht: Deskriptive Analyse der Daten mit einer geografische Zuordnung zu einzelnen Regionen

# *Zusammenfassung*

- Systematische Vorgangsweise bei datengestützten Fragestellungen ist von enormer Wichtigkeit
- Daten ≠ Informationen
- Datenanalyse führt meist zur Datenreduktion
  - Einsatz von statistischen Methoden zwingend notwendig
  - Führt meist von Big Data zu Smart Data
  - Basis für Verbesserungen
- Datenanalyse kann in folgenden Bereichen eingesetzt werden
  - Identifikation von Zusammenhängen – statistische Modellierung
  - Monitoring
  - Prognose
- „Mache die Dinge so einfach wie möglich – aber nicht einfacher“. (Albert Einstein)

# *Resümee und Ausblick*

- Datenanalyse sollte integraler Bestandteil im unternehmerischen Umfeld sein
- Sammeln und auswerten der „richtigen“ Daten wird zum Erfolgsfaktor
- Wissen aus den Daten wird Differenzierungsmerkmal am Markt

Angewandte Statistik ist eine Zusammenfassung von Methoden, die uns erlauben, vernünftige Entscheidungen im Falle von Ungewissheit zu treffen.

(Abraham Wald)

# Danke für Ihre Aufmerksamkeit!

*DI Hermann Katz*

*JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH*

*POLICIES – Institut für Wirtschafts- und Innovationsforschung  
Datenanalyse und statistische Modellierung*

*Leonhardstraße 59  
8010 Graz*

*Tel.: +43 316 876-1553*

*Mobil: +43 664 602 876 1553*

*PC-Fax: +43 316 8769-1553*

*Email: [hermann.katz@joanneum.at](mailto:hermann.katz@joanneum.at)*



Der DIH SÜD wird  
unterstützt von:



LAND  KÄRNTEN