

DIH SÜD - DIGITAL INNOVATION HUB SÜD

Deskriptive Datenanalyse Teil 2

Anwendung und Umsetzung

16.03.2023



Der DIH SÜD wird
unterstützt von:



LAND  KÄRNTEN

Deskriptive Datenanalyse

Anwendung und Umsetzung

HERMANN KATZ

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

JOANNEUM RESEARCH Forschungsgesellschaft mbH

hermann.katz@joanneum.at

Graz, 16. März 2023

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 – 15:00: Zusammenfassung und Resümee



VORGANGSWEISE BEI DATENGESTÜTZTEN FRAGESTELLUNGEN

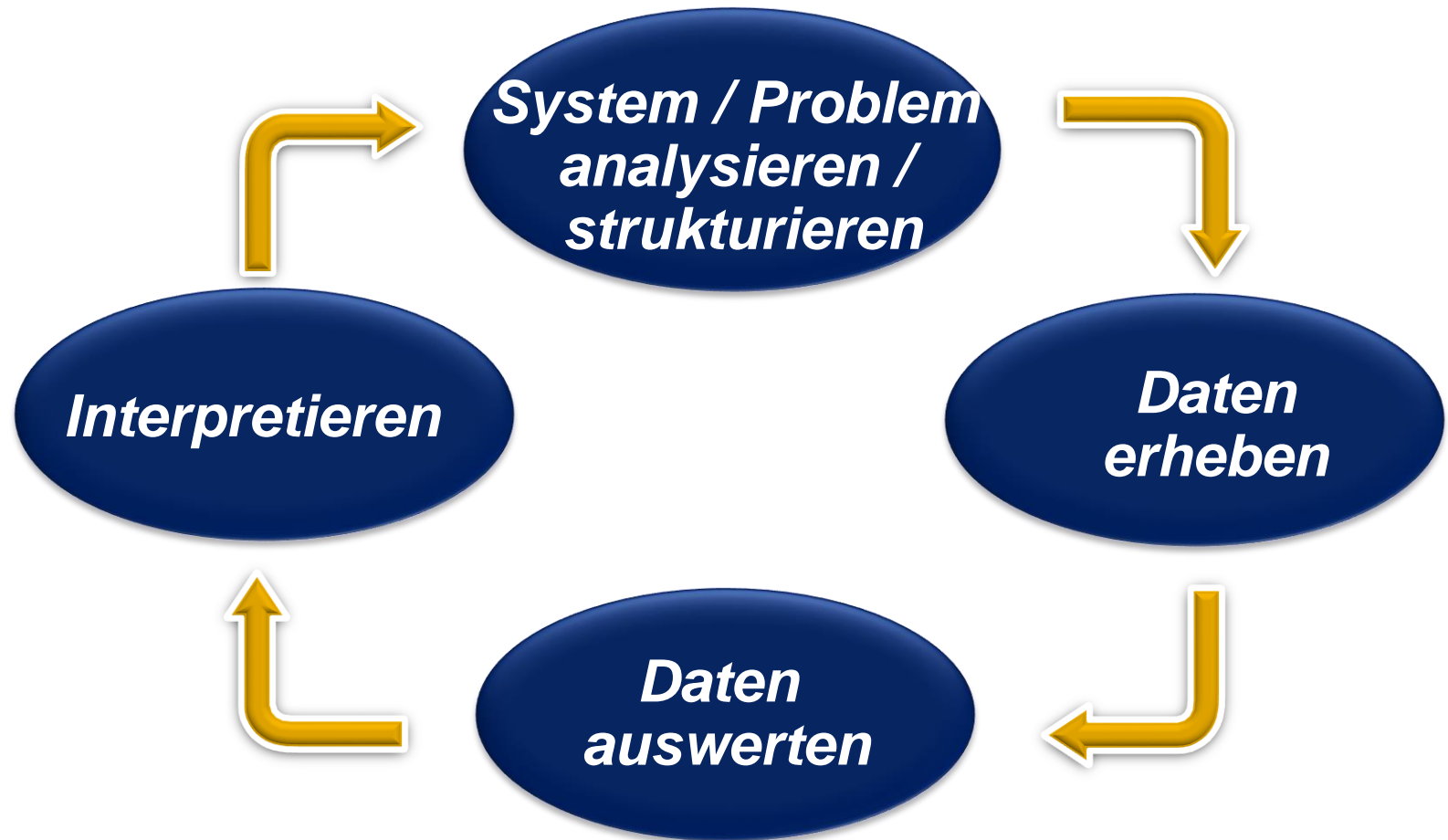


Der DIH SÜD wird
unterstützt von:



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 Stichprobenumfanges
 - 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

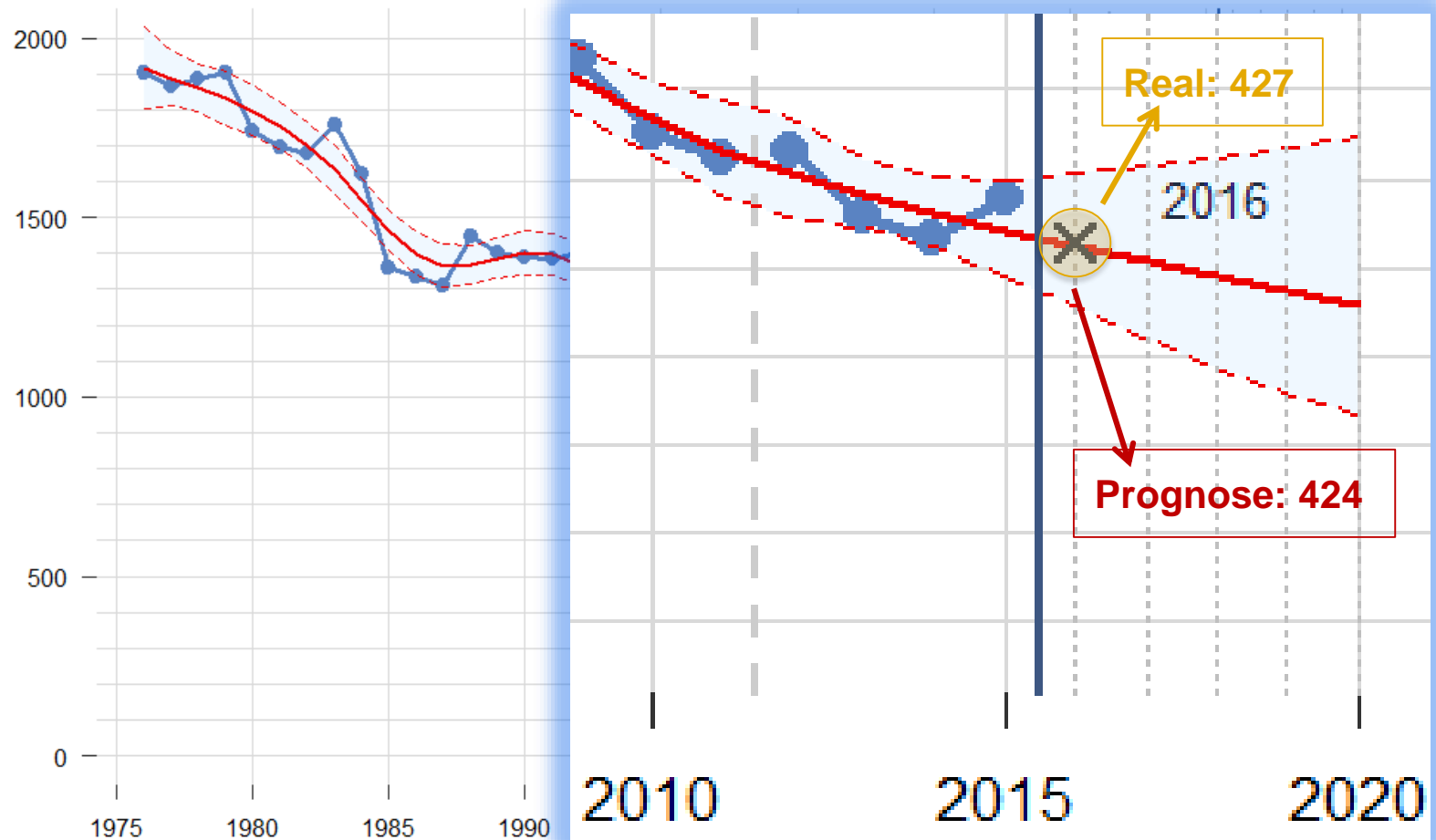
Model the Data

Vorhersagen in Zukunft

Communicate and visualize the result

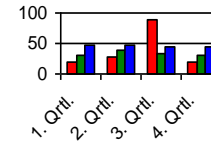
Bericht

KFV Standard Reporting - Prognose der Todesfälle im Straßenverkehr

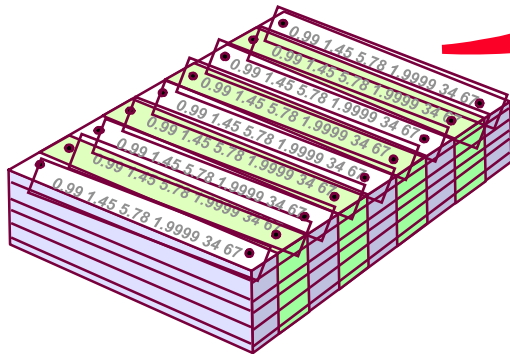


Daten \neq Information

Information

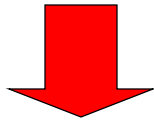


Statistische Werkzeuge



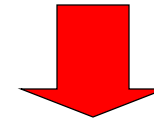
Gliederung der Statistik

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

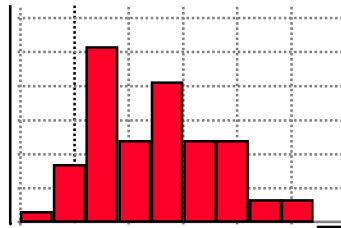
■ Methoden:

- Grafische Darstellungen
- Kennzahlen (Maßzahlen)

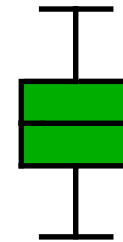
Unterschiedliche Grafiken



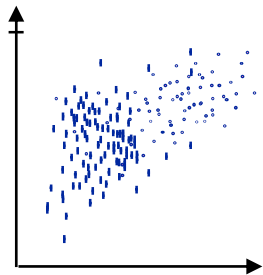
Piechart



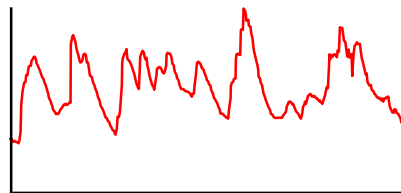
Histogramm



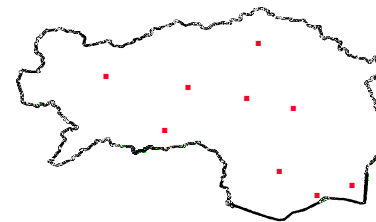
Boxplot



Scatterplot



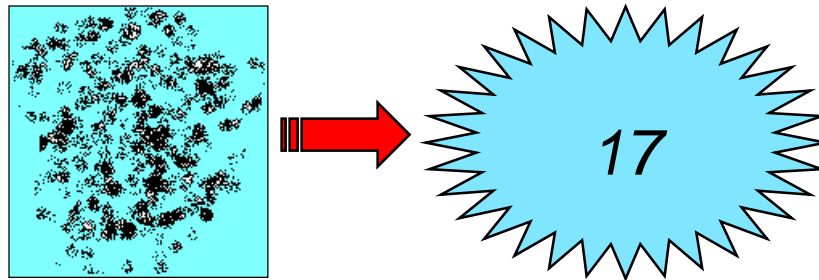
Zeitreihe



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

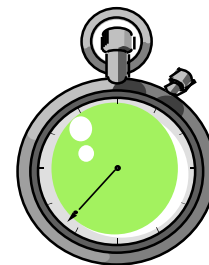
Skalentyp	Ordinalskala
Definierte Relationen und Operationen	$\neq = < >$
Zulässige Transformationen	isotone oder rangerhaltende Transformationen
Beispiele für Merkmale	Zeugnisnoten, Sozialstatus, Produktgüteklassen, Mercalli-Erdbebenskala
Merkmalsausprägung	Ordinalzahlen (i.d.R. ganze Zahlen)

gut
besser
G

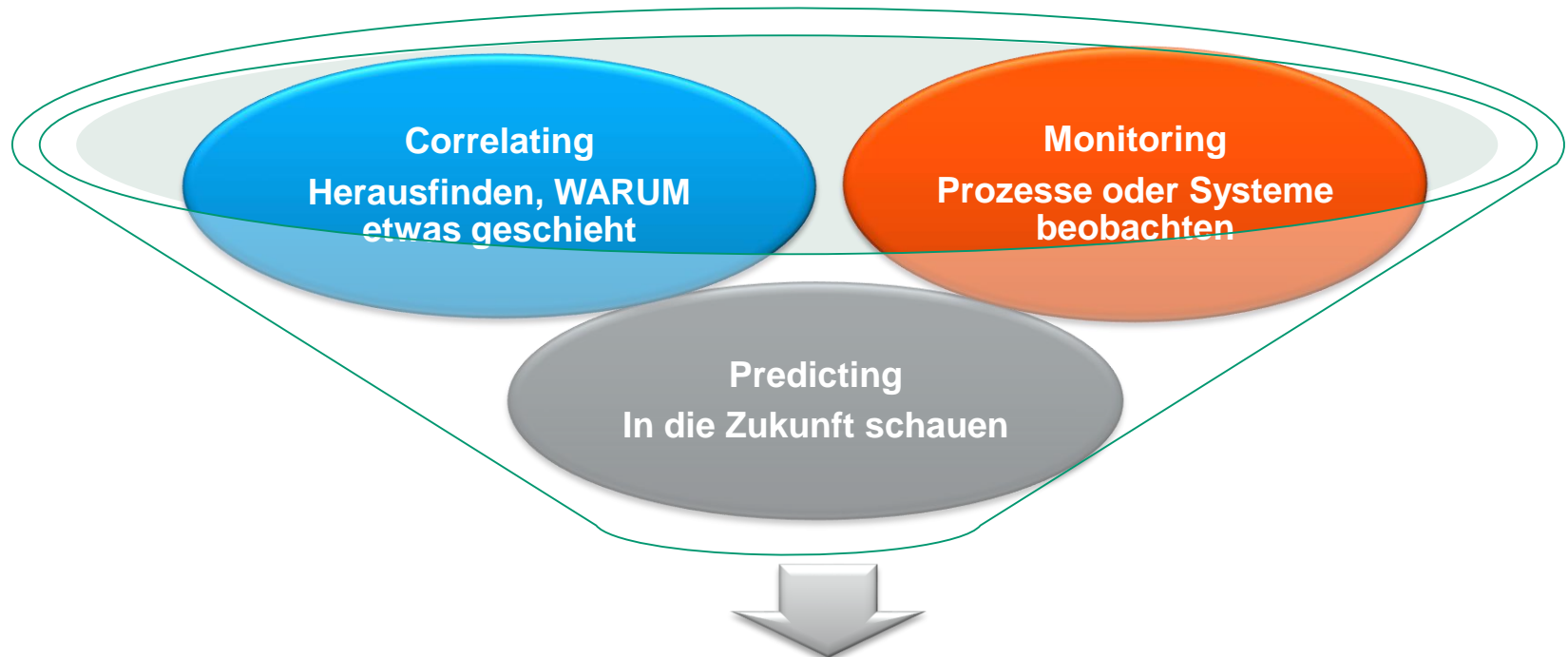


Skalierung

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

23

- biosanica GmbH ist ein Lebensmittel Großhändler mit Geschäftssitz in Süderholz, Deutschland
- Spezialisiert auf (Bio)-Trockenäpfel in jeder Form
 - Apfelstücke, -chips, -würfel und -pulver
- Forschungsfragen
 - Welche sinnvollen Kennzahlen/Werte kann man aus den Daten ermitteln und diese in einem Dashboard grafisch darstellen?
 - Wie können die Daten besser/übersichtlicher strukturiert werden?
 - Durchführung einer ABC-Analyse nach Kunden bzw. Artikeln
- Für alle drei Forschungsfragen wird Unterscheidung zwischen Bio-Anbau und konventionellen Anbau bzw. eine gesamtheitliche Betrachtung durchgeführt

Datensituation

24

- Datensatz beinhaltet Vertriebsplanung für ein Wirtschaftsjahr
 - Wirtschaftsjahr beginnt mit September (Apfelernte)
 - Detaillierte Auflistung von Planabsatzmenge und Planumsatz unterteilt nach Kunden und Artikelnummer
 - Monatliches Update der Planabsatzmenge und des Planumsatzes (V1 – V6)
- Zusätzlich beinhaltet Datensatz auch die tatsächliche Absatzmenge und den Umsatz für einzelne Monate, sowie zugehörige Planwerte fürs Monat (V1 - V6)
 - Auch hier detaillierte Auflistung der tatsächlichen Menge/Umsatz nach Kunden und Artikelnummer

Idee

25

- Erstellung eines Dashboards mittels einer Shiny-App
- Shiny ist ein Paket der Open Source Software R
- Vorteile von Shiny:
 - Mittels Shiny können Webanwendungen einfach erstellt werden da vorgefertigte Ein- und Ausgabe-Bedienelemente vorinstalliert sind
 - Interaktive Benutzung der App und live Anpassungen vom Benutzer möglich
 - Anwender kann ohne Hintergrundwissen zur Programmiersprache oder mathematischen Methoden die notwendigen Tools verwenden

Ergebnisse

D:/wed/DIH-Kurs/02_work/Biosanica_App/Biosanica_App/startApp_developmode.R - Shiny

http://127.0.0.1:7388 | Open in Browser

Publish

biosanica GmbH

START

Mengen & Umsatzvergleich

Artikeldarstellung

Artikeldarstellung nach Kunden

ABC-Analyse

Kontrakterfüllung

IST

PLAN

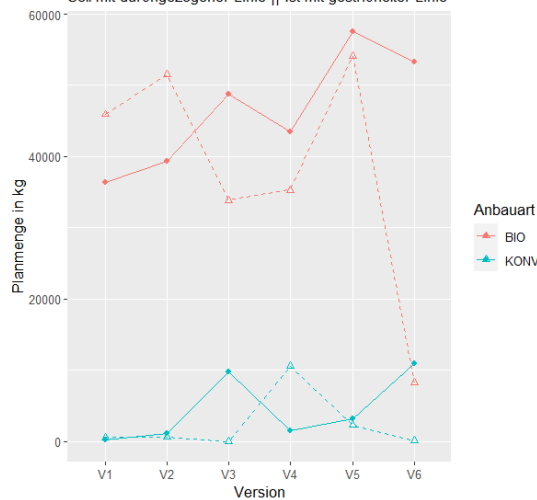
Auswahl

Menge

Umsatz

Vergleich zwischen Soll- und Istmenge

Soll mit durchgezogener Linie || Ist mit gestrichelter Linie



Zeitlicher Verlauf der Differenz

Show 15 entries

Search:

	Version	Anbauart	Sollmenge	Istmenge	Differenz
1	V1	BIO	36320	45909.15	95
2	V1	KONV	200	580	
3	V2	BIO	39370	51489.22	121
4	V2	KONV	1080	607.5	
5	V3	BIO	48820	33860	-1
6	V3	KONV	9800	25	
7	V4	BIO	43450	35279.7	-8
8	V4	KONV	1450	10600	
9	V5	BIO	57100	54100	

Präsentation Use Case 2



Der DIH SÜD wird
unterstützt von:



LAND  KÄRNTEN

Systemanalyse / 1

- Nuki Home Solutions GmbH steht für nachrüstbare Zutrittslösungen
- Nuki ist ein schnell wachsendes Tech Unternehmen aus Graz, mit Kunden aus aller Welt. Das Hauptprodukt ist das Nuki Smart Lock, welches ein elektromagnetisches Schloss ist, das es den Kunden ermöglicht, ohne Schlüssel, entweder komplett automatisch bei Annäherung oder durch Zubehör (z.B. Fingerabdrucksensor) die Wohnung bzw. ihr Haus zu betreten. Die Kunden verwenden das Produkte zum Teil unter sehr unterschiedlichen Bedingungen was einerseits Umweltfaktoren (Temperatur, Feuchtigkeit,...) und auch die Nutzungsfrequenz (Anzahl Sperrvorgänge) betrifft

Systemanalyse / 2

- Es gibt eine Fehlerkategorie (2 Fehler intern bezeichnet als 4A und 4B), deren genaue Ursache noch ungeklärt ist.
- Die Techniker vermuten, dass es sich dabei um eine Kombination von verschiedenen Faktoren - ziemlich sicher u.a. der Sperrzyklen, eventuell aber auch Temperatur, Luftfeuchtigkeit, eventuell sogar Parameter wie verwendetes Zubehör, Türtyp, Verbrauchszahlen etc. mehr oder weniger starken Einfluss auf die Produktlebenszeit haben.
- Fragestellung: Können aus den zur Verfügung gestellten Daten (Land, Wetter (Außentemperatur und Luftfeuchtigkeit), Sperrzyklen, Produktionsfirma und Produktionsdatum (Wochentag/Uhrzeit) konkrete Einflüsse auf die Produktlebenszeit im Zusammenhang mit dem Fehler 4A oder 4B getroffen werden?

Datenanalyse

- Modellierungsansatz
 - Zielgröße: Aktion bzw. zugehöriger Status
 - Status ist Ergebnis der Aktion
 - erfolgreich abgeschlossen/geöffnet
 - abgebrochen
 - Fehler Motor blockiert - Fehler 4A
 - Kupplungsfehler - Fehler 4B
 - Einflussgrößen sind alle relevanten Daten zum Schloss (Stamm-, Konfigurations- und Protokolldaten, Umgebungsdaten)
 - Qualität der Daten ist der Erfolgsfaktor bei dieser datengestützten Fragestellung!

Ergebnisse

- Identifikation jener Einflussgrößen, die auf den Schließvorgang einen positiven/negativen Einfluss haben
- Modellierung bezogen auf Fehler unabhängig vom Fehlertyp
- Modellierung nach Fehlertyp 4A und AB
- Ergebnis: Identifikation jener Rahmenbedingungen, die für einen Schließvorgang förderlich/nicht förderlich sind!
- Handlungsempfehlungen für den Kunden, um diese Fehlertypen zu vermeiden!

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



Der DIH SÜD wird
unterstützt von:



LAND  KÄRNTEN