|  |
| --- |
| Schriftliche Ausarbeitung |
| Case Study 1 SE S2020  Till Bieg, David Krug – Fachhochschule Wiener Neustadt |

INHALT

[Schriftliche Ausarbeitung 1](#_Toc43239036)

[Tabellenverzeichnis 2](#_Toc43239037)

[Abbildungsverzeichnis 2](#_Toc43239038)

[1 Beschreibung des Technologie-Stacks 3](#_Toc43239039)

[2 Beschreibung des ETL-ProzessES 3](#_Toc43239040)

[3 Beschreibung der Datenbasis Und Prüfung der Integrität 4](#_Toc43239041)

[3.1 Überblick 4](#_Toc43239042)

[3.2 Bemerkungen zur Vollständigkeit der Datenbasis 4](#_Toc43239043)

[3.3 Grundlegende Überprüfung der Datenintegrität 4](#_Toc43239044)

[4 Weiterführende Analysen 5](#_Toc43239045)

[5 Schlussfolgerungen und Empfehlungen 7](#_Toc43239046)

[5.1 … 7](#_Toc43239047)

[5.1.1 … 7](#_Toc43239048)

[6 Referenzen 8](#_Toc43239049)

# Tabellenverzeichnis

[Tabelle 1. Buchungseinträge mit unvollständigen Zeilennummern. 4](#_Toc43239035)

# Abbildungsverzeichnis

[Abbildung 1. Visualisierung des ELT-Prozesses. 3](#_Toc43239029)

[Abbildung 2. Verteilung der Buchungshäufigkeit pro Benutzer. 5](#_Toc43239030)

[Abbildung 3. Anzahl an Buchungen pro Department. 6](#_Toc43239031)

[Abbildung 4. Anzahl an Buchungen pro Monat. 6](#_Toc43239032)

[Abbildung 5. Anzahl an Buchungen pro Wochentag (Belegdatum). 6](#_Toc43239033)

[Abbildung 6. Anzahl der Buchungen pro Wochentag (Erfassungsdatum). 7](#_Toc43239034)

# Beschreibung des Technologie-Stacks

Um die gegebenen Daten zu analysieren, wurde

* Microsoft Excel 2019 (MS Excel)
* R, Version 3.6.2 mit RStudio und folgenden Packages (Auswahl):
  + readxl: Zum Import von MS Excel-Dateien in R
  + dplyr: Zur Datenmanipulation (z.B. Filtern, Selektieren, Aggregieren)
  + tidyr: Zur Datenmanipulation (z.B. Pivotieren)
  + lubridate: Zur Manipulation von zeitbezogenen Variablen (z.B. Datum)
  + ggplot2: Zur Erstellung statischer Grafiken
  + plotly: Zur Erstellung von interaktiven Grafiken
  + shiny: Zur Erstellung von Dashboards und interaktiven Grafiken
  + rmarkdown: Zur Erstellung und Export von Analyseberichten
  + benford.analysis: Zur Durchführung von Benford-Analysen
  + tsoutliers: Zur Analyse von Zeitreihen-Daten (z.B. Ausreißerdetektion)

Je nach Anforderungen, die im Verlauf der Analyse erkennbar werden, behalten wir uns vor, zusätzliche Packages zu verwenden, um die Analyse nach besten Möglichkeiten durchzuführen.

# Beschreibung des ETL-ProzessES

Die Daten aus den bereitgestellten MS Excel-Dateien wurden unter Verwendung des readxl-Packages unmittelbar in die globale Umgebung von R geladen. Hierbei ist darauf zu achten, dass die Daten innerhalb einer MS Excel-Datei gegebenenfalls auf mehrere Arbeitsblätter aufgegliedert sind. Obwohl die Datenmenge insgesamt relativ umfassend ist, wird es ohne Weiteres möglich sein, diese Daten in den Arbeitsspeicher zu laden und entsprechende Analysen durchzuführen. Deswegen ist die Verwendung einer Datenbank im Rahmen des ETL-Prozesses nicht unbedingt notwendig.

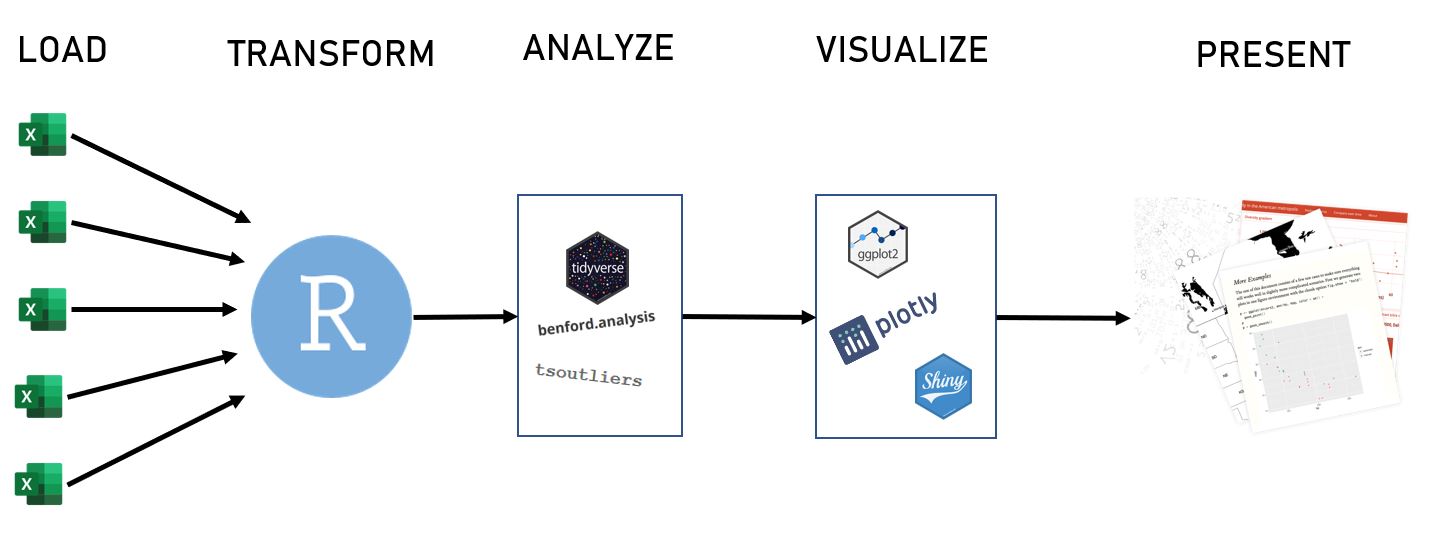


Abbildung 1. Visualisierung des ELT-Prozesses.

Für die Datenaufbereitung und -analyse wird wie bereits erwähnt die statistische Programmiersprache R, Version 3.6.2, verwendet (R Core Team, 2019). Dabei bedienen wir uns der umfangreichen Auswahl an Packages des tidyverse (z.B. dyplr, tidyr, lubridate). Für spezifische Analysen (zum Beispiel Benford- oder Zeitreihen-Analyse) werden darüber hinaus weitere Packages verwendet (zum Beispiel benford.analysis und tsoutliers). Um die Daten zu explorieren und Ergebnisse anschaulich darzustellen wird ein wesentlicher Schwerpunkt auch auf der Datenvisualisierung liegen. Diesbezüglich ist die Verwendung von Packages wie ggplot2, shiny und plotly angedacht.

Für die Vorbereitung der Präsentation und die schriftliche Ausarbeitung werden wir in erster Linie rmarkdown verwenden, wobei ein Rückgriff auf andere Formate (z.B. shiny) nicht im Vorhinein ausgeschlossen ist.

# Beschreibung der Datenbasis Und Prüfung der Integrität

## Überblick

Wie bereits im Konzept beschrieben, liegt der Schwerpunkt der Analyse auf der Prüfung der Daten auf dolose Handlungen / betrügerische Aktivitäten. Folgende Datensätze wurden von der ABC Gesellschaft für die Prüfung zur Verfügung gestellt.

* Buchungsjournal für das Jahr 2014: Diese Datei enthält die Buchungseinträge aus dem Jahr 2014 (entspricht der zu prüfenden Periode).
* Buchungsjournal für das Jahr 2013: Diese Datei enthält die Buchungseinträge aus dem Jahr 2013 (Vorjahr der zu prüfenden Periode).
* Saldenliste („Trial Balances“) für das Jahr 2014: Diese Datei die Anfangs- und Endsalden jedes Kontos für das Jahr 2014.
* Saldenliste für das Jahr 2014 nur mit Daten von Jänner bis September: Geeignet für Forecasts.
* Saldenliste für das Jahr 2013: Beinhaltet die Anfangs- und Endsalden jedes Kontos für das Jahr 2013.
* Kontenplan („Chart of Accounts“): Der Kontenplan beinhaltet die Gliederung jedes Konto zu Bilanz- oder GuV-Posten.
* Quellenliste („Source Listing“): Der Inhalt der Quellenliste ist nicht ganz schlüssig, da es scheint, als ob viele Einträge bloße Platzhalter enthalten. Ist mit der ABC Gesellschaft abzuklären.
* Benutzerliste (User Listing): Enthält eine Zuordnung verschiedener Systembenutzer zu Benutzernamen und Departments.

## Bemerkungen zur Vollständigkeit der Datenbasis

In Bezug auf die Vollständigkeit sind vorab einige Bemerkungen zu machen. Das Buchungsjournal enthält in Bezug auf die meisten Variablen vollständige Angaben. Insbesondere enthalten die Spalten zum funktionalen Betrag („Functional Amount“) und der Benutzer-ID (Preparer ID) keine fehlende Werte.

Außerdem ist anzumerken, dass die Datengrundlage Unvollständigkeiten aufweist, die die Prüfbarkeit einschränken (fehlenden Buchungsbeschreibungen, fehlende Buchungsuhrzeit).

## Grundlegende Überprüfung der Datenintegrität

Alle im Folgenden vorgestellten Überprüfungen der Datenintegrität lassen sich anhand der beigefügten Datei *„analysis.rmd“* nachvollziehen.

Zunächst wurde überprüft, ob die Summe aller funktionalen Beträge für das Jahr 2014 0 entspricht. Dieses Integritätskriterium konnte bestätigt werden. Auch die Summen innerhalb jedes Journaleintrags (Summe der Buchungszeilen pro Journaleintrag) entsprechen jeweils, so dass auch dieses Kriterium erfüllt ist. Darüber hinaus wurde geprüft, ob die jeweiligen Anfangs- und Endsalden pro Konto zwischen dem Buchungsjournal und der Saldenliste überstimmen. Auch das konnte für alle Konten bestätigt werden. Allerdings zeigte eine weitere Prüfung, dass die **Zeilennummern in Bezug auf die jeweiligen Journaleinträge nicht immer lückenlos** sind. Es gibt zumindest 9 Journaleinträge, die nicht lückenlos sind. Diese Journaleinträge sind in Tabelle 1 dargestellt.

Tabelle 1. Buchungseinträge mit unvollständigen Zeilennummern.

|  |  |  |
| --- | --- | --- |
| **JE-Nummer** | **Tatsächliche Anzahl an Zeilen** | **Erwartete Anzahl an Zeilen** |
| 2014-0001-2400008088 | 4 | 8 |
| 2014-0001-2400013762 | 4 | 8 |
| 2014-0001-4200036317 | 2 | 3 |
| 2014-0001-4200036320 | 2 | 3 |
| 2014-0001-4200036323 | 2 | 3 |
| 2014-0001-4200036332 | 2 | 3 |
| 2014-0001-4200039194 | 4 | 7 |
| 2014-0001-4200046717 | 2 | 3 |
| 2014-0001-4200050312 | 2 | 3 |

Wie in der Tabelle zu sehen, fehlen je nach Eintrag bis zu vier Buchungszeilen. Das kann ein erster Hinweis auf dolose Handlungen o.ä. sein. Die Buchungen sind assoziiert mit insgesamt zwei individuellen Benutzern *(„SYS“ und „SheAl01“),* zwei Quellengruppen *("Accruals" und "Goods receipts")* und vier verschiedenen Konten *("0060000500", "0140520850", "0140203005", "0140203004").* Diese lückenhaften Journaleinträge werden im Rahmen der weiterführenden Analyse noch genauer betrachtet.

Zusammenfassend erfüllen die zur Verfügung gestellten Daten die überwiegende Anzahl grundlegende Integritätskriterien. Allerdings gibt es Journaleinträge mit unvollständigen Zeilennummern, was ein erster Hinweis auf dolose Handlungen ist und bei den weiterführenden Analysen genauer betrachtet wird.

# Weiterführende Analysen

## Prüfungen der buchenden Personen

…

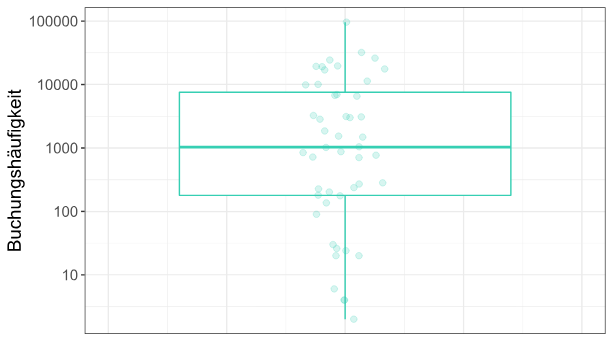


Abbildung 2. Verteilung der Buchungshäufigkeit pro Benutzer.

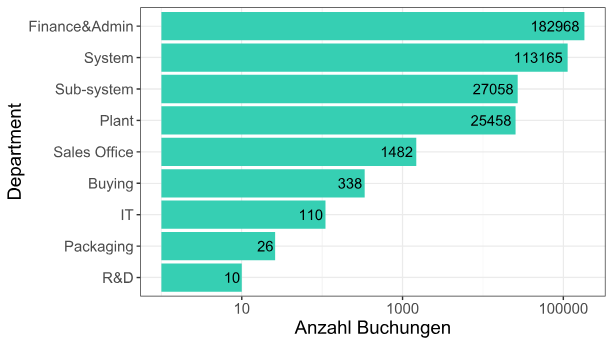


Abbildung 3. Anzahl an Buchungen pro Department.

## Prüfungen der Buchungs- und Erfassungszeit

…

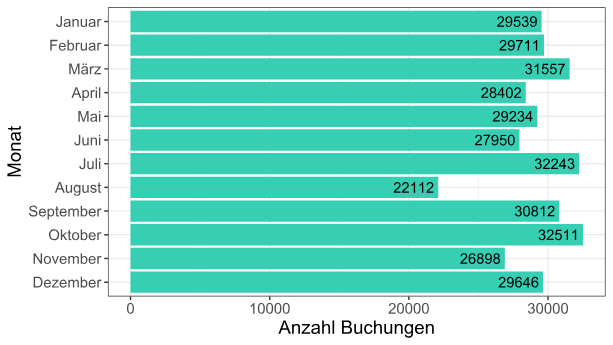


Abbildung 4. Anzahl an Buchungen pro Monat.

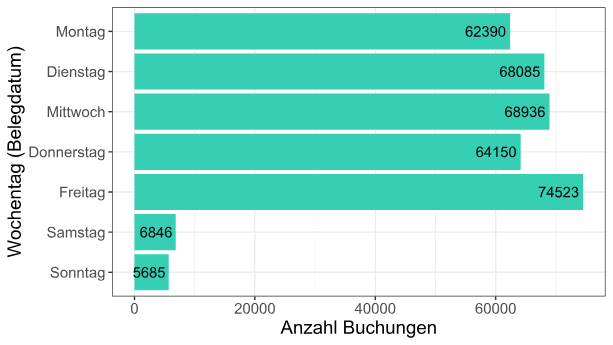


Abbildung 5. Anzahl an Buchungen pro Wochentag (Belegdatum).

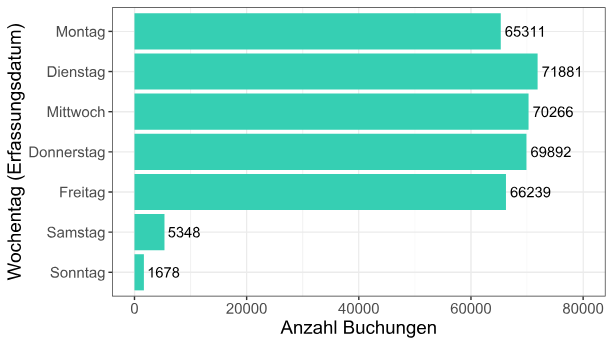


Abbildung 6. Anzahl der Buchungen pro Wochentag (Erfassungsdatum).

## Prüfung der verwendeten Konten

…

## Prüfung der Höhe beziehungsweise Ziffern und Beträgen

…

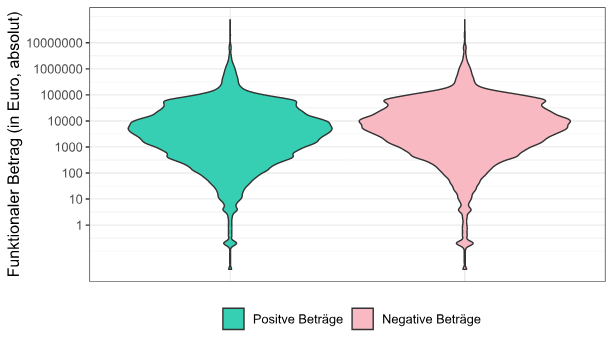


Abbildung 7. Verteilung der funktionalen Beträge nach positiven und negativen Beträgen.

…

## Prüfungen auf Unterschiede zum Vorjahr (Trendanalysen)

…

# Schlussfolgerungen und Empfehlungen

Hinweise auf dolose Handlungen / Verletzungen der Integrität:

* Einträge mit vollständigen Zeilennummern
* Einträge mit Eintragsdatum vor Belegdatum
* Einträge mit Beleg- und/oder Erfassungsdatum am Wochenende

## …

### …

# Referenzen

Nigrini, M. J., & Miller, S. J. (2009). Data diagnostics using second-order tests of Benford's law. Auditing: A Journal of Practice & Theory, 28(2), 305-324.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.