|  |
| --- |
| 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)

Diese Ausarbeitung wurde im Rahmen der Lehrveranstaltung Case Study 1 ILV S2020 an der Fachhochschule Wiener Neustadt erstellt. Sie beschreibt den Technology-Stack, ETL-Prozess und durchgeführte Analysen hinsichtlich eines rechnungslegungsbezogenen Datensatzes, der innerhalb der Lehrveranstaltung zur Verfügung gestellt wurde. Der Schwerpunkt der Analysen liegt dabei auf der Untersuchung möglicher betrügerischer Aktivitäten. Die Ergebnisse der Analysen werden abschließend in Form von Schlussfolgerungen zusammengefasst.

# Beschreibung des Technologie-Stacks

Um die gegebenen Daten zu analysieren, wurden folgende Softwaretools verwendet:

* *Microsoft Excel 2019* (*MS Excel*)
* *R*, Version 3.6.2 (R Core Team, 2019), mit *RStudio* und folgenden Packages:
  + *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
  + *ggforce*: Zur Erstellung statischer Grafiken
  + plotly: Zur Erstellung interaktiver Grafiken
  + *rmarkdown*: Zur Erstellung und Export von Analyseberichten
  + *benford.analysis*: Zur Durchführung von Benford-Analysen
  + *tsoutliers*: Zur Analyse von Zeitreihen-Daten (z.B. Ausreißerdetektion)
  + *naniar*: Zur Analyse fehlender Werte
  + *janitor*: Zur Datenbereinigung
  + *here*: Zum Management von Systempfaden

Zur Erstellung der finalen Ausarbeitung und Präsentation wurden *Microsoft Word 2019* (MS Word) und *Microsoft Powerpoint 2019 (MS Powerpoint)* verwendet. Außerdem wurde aufgrund des Schwerpunkts der Analyse auf betrügerischen Aktivitäten und dolosen Handlungen auf die Erstellung eines Dashboards (z.B. mittels *shiny*) verzichtet. Ein Dashboard ist vor allem zur Darstellung von Kennzahlenanalysen sehr gut geeignet – zur die detaillierten Darstellung der Betrugsanalysen erscheint es aber eher unflexibel. Für die Präsentation wird stattdessen eine R-Session vorbereitet, innerhalb derer nach Bedarf Detailanalysen durchgeführt werden.

# Beschreibung des ETL-ProzessES

Für die Datenaufbereitung und -analyse wurde wie bereits erwähnt die statistische Programmiersprache *R*, Version 3.6.2, verwendet (R Core Team, 2019).

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

Zur weiteren Analyse wurde auf eine umfangreiche Auswahl an verfügbaren Packages zurückgegriffen (z.B *dyplr, tidyr, lubridate,* siehe Abschnitt 1). Zur visuellen Exploration der Daten und anschaulichen Darstellung der Analyse wurden Packages zur Datenvisualisierung wie *ggplot2* verwendet. Spezifische Analysen wie Benford- oder Zeitreihen-Analyse wurden unter Verwendung weiterer, dezidierter Packages durchgeführt (zum Beispiel *benford.analysis* und *tsoutliers*).

Zur Dokumentation der Analyse sowie Vorbereitung von Präsentation und schriftlicher Ausarbeitung wurde das Package *rmarkdown* verwenden. Die damit erstellte Analysedokumentation wurden nachfolgend in .*docx*- und *pptx*-Formate exportiert und mit *MS Word* und *MS Powerpoint* weiter angepasst. Abbildung 1 stellt den ETL-Prozess visuell dar.

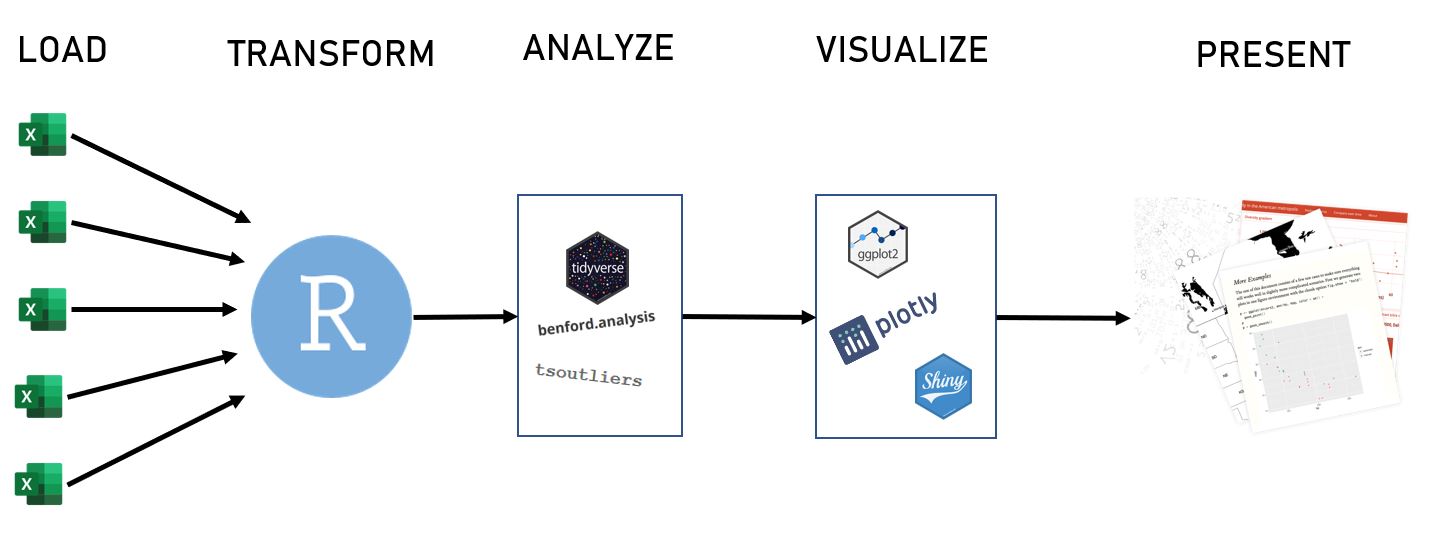


Abbildung 1. Visualisierung des ETL-Prozesses.

# 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.
* 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.
* Belegliste („Source Listing“): Enthält eine Zuordnung von Belegen zu Belegbeschreibungen und Beleggruppen.
* Benutzerliste („User Listing“): Enthält eine Zuordnung verschiedener Systembenutzer zu Benutzernamen und Departments.

## Bemerkungen zur Vollständigkeit der Datenbasis

Das Buchungsjournal für die zu prüfende Periode 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. Problematisch sind die fehlenden Werte in Bezug auf die Variable „JE Line Description“: Diese Variable enthält anstatt des Buchungstexts (Beschreibung der Buchung) ausschließlich Platzhalter oder fehlende Werte. Das schränkt die Prüfbarkeit der Daten ein, da der Buchungstext unter Umständen Rückschlüsse auf den Inhalt der jeweiligen Buchung geben kann. Des Weiteren ist anzumerken, dass die Uhrzeit der Buchung beziehungsweise Erfassung zur Verfügung gestellt wurde, was eine weitere Limitation bezüglich der Prüfbarkeit darstellt (die Buchungszeit könnte beispielsweise genutzt werden, um Buchungen[[1]](#footnote-1) außerhalb der Geschäftszeiten zu identifizieren). Des Weiteren enthält der Kontenplan 15.1% fehlende Werte in Bezug auf den Kontonamen und die Kontoklassen, so dass eine nähere Beschreibung und Eingliederung für diese Konten fehlt. Weitgehend unbrauchbar für eine Prüfung erscheint die Belegliste, da diese eine hohe Anzahl von Platzhaltern (bis zu 74.4%) anstelle von aussagekräftigen Beschreibungen enthält. In der Benutzerliste sind Rolle und Titel des jeweiligen Mitarbeiters nicht enthalten. Diese Variablen könnten im Rahmen der Prüfung ebenfalls wertvolle Rückschlüsse erlauben. Die Daten der zur Verfügung gestellten Saldenlisten sind so weit vollständig.

Zusammenfassend sind die zu Verfügung gestellt Daten weitgehend vollständig – allerdings schränken einige fehlende Werte in den verschiedenen Datensätzen die Prüfbarkeit ein.

## Grundlegende Überprüfung der Datenintegrität

Die 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 Beleggruppen *("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 einen überwiegenden Teil der grundlegenden Integritätskriterien. Allerdings gibt es Journaleinträge mit unvollständigen Zeilennummern, was ein erster Hinweis auf dolose Handlungen ist und bei den weiterführenden Analysen genauer untersucht wird.

# Weiterführende Analysen

## Prüfungen der buchenden Personen

Die folgenden Analysen beziehen sich schwerpunktmäßig auf die buchenden Personen, um mögliche Auffälligkeiten in diesem Kontext zu identifizieren und genauer zu untersuchen. Im Buchungsjournal haben insgesamt 48 Benutzer Buchungen durchgeführt. Abbildung 2 stellt die Verteilung der Buchungshäufigkeit nach Benutzern dar (die Buchungshäufigkeit ist dabei auf einer logarithmischen Skala visualisiert). Wie in der Abbildung zu erkennen gibt es erhebliche Unterschiede in der Buchungshäufigkeit zwischen Nutzern: So gibt es einen Nutzer mit fast 100000 Buchungen, andere Benutzer haben weniger als 10 Buchungen insgesamt. Der Median der Buchungshäufigkeit liegt bei ca. 1000 Buchungen, die Spannweite ist allerdings erheblich. Die Buchungshäufigkeit kann als Indikator für die Expertise von Benutzern sein – dahinter steht beispielsweise die Annahme, dass Mitarbeiter, die häufiger Buchungen durchführen und daher weniger Fehler machen.

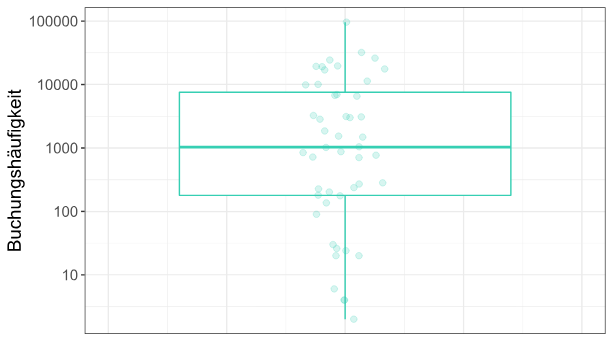


Abbildung 2. Verteilung der Buchungshäufigkeit pro Benutzer.

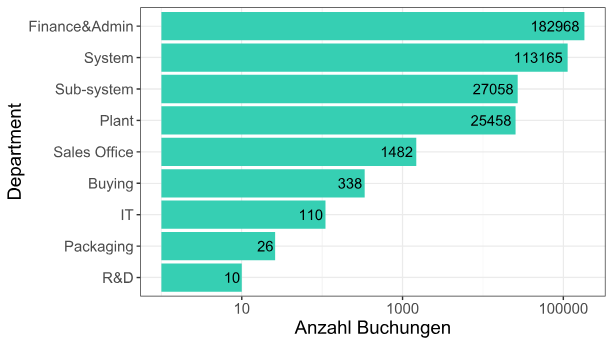


Abbildung 3. Anzahl an Buchungen pro Department.

In diesem Zusammenhang erscheint auch ein Blick auf die Buchungshäufigkeiten der verschiedenen Departments sinnvoll, denen die verschiedenen Benutzer zugeordnet sind. Das gibt Aufschluss darüber, welche Departments typischerweise Buchungen durchführen. Abbildung 3 stellt die Anzahl an Buchungen für die verschiedenen Departments dar. Auch zwischen den Departments zeigen sich erhebliche Unterschiede bei den Buchungshäufigkeiten: Das Department „Finance&Admin“ hat über 180000 Buchungszeilen erfasst, das Department „R&D“ nur insgesamt 10 Buchungen.

Im Folgenden sollen darüber hinaus jene zwei Benutzer genauer betrachtet werden, die mit den lückenhaften Journaleinträgen assoziiert sind. Tabelle 2 stellt Merkmale dieser Benutzer da.

Tabelle 2. Merkmale der Benutzer, die mit lückenhaften Buchungszeilen assoziiert sind.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Benutzername** | **Ganzer Name** | **Department** | **Rolle** | **Titel** | | **Anzahl Buchungen** |
| SheAl01 | Alan Shepard | Finance&Admin | *Fehlend* | | *Fehlend* | 17561 |
| SYS | SYS SYS | System | *Fehlend* | | *Fehlend* | 16952 |

Beide Benutzer gehören Departments an, die viele Buchungen durchgeführt haben. Die Benutzer selbst haben auch überdurchschnittlich viele Buchungen durchgeführt. Das lässt darauf schließen, dass es sich um „erfahrene“ Benutzer, die vergleichsweise wenige Fehler bei der Systembenutzung machen. Es erscheint also tendenziell unwahrscheinlich, dass die lückenhaften Journaleinträge durch Irrtum zustandegekommen. Es empfiehlt sich jedenfalls weitere Details in Bezug auf diese Benutzer mit dem Management der ABC-Gesellschaft zu besprechen.

## Prüfungen der Buchungs- und Erfassungszeit

…

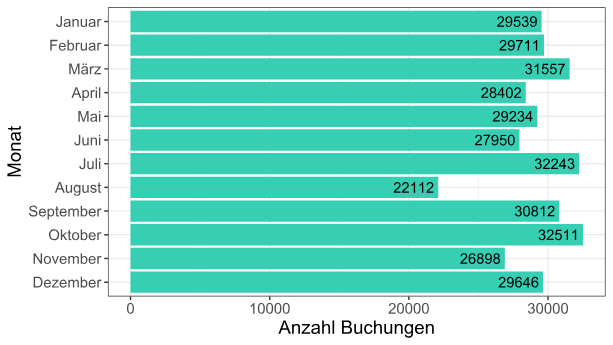


Abbildung 4. Anzahl an Buchungen pro Monat.

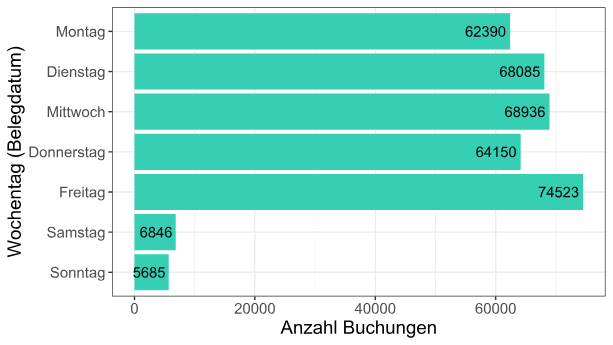


Abbildung 5. Anzahl an Buchungen pro Wochentag (Belegdatum).

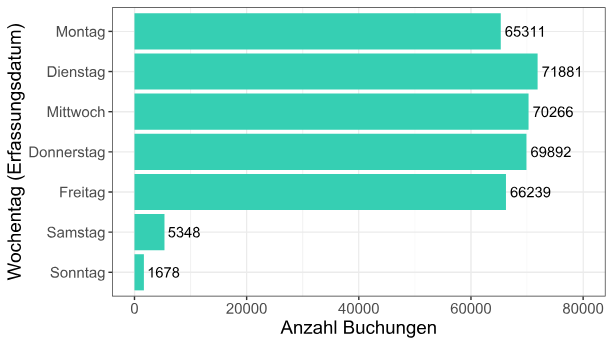


Abbildung 6. Anzahl der Buchungen pro Wochentag (Erfassungsdatum).

## Prüfung der verwendeten Konten

…

## Prüfung der Höhe beziehungsweise Ziffern und Beträgen

…

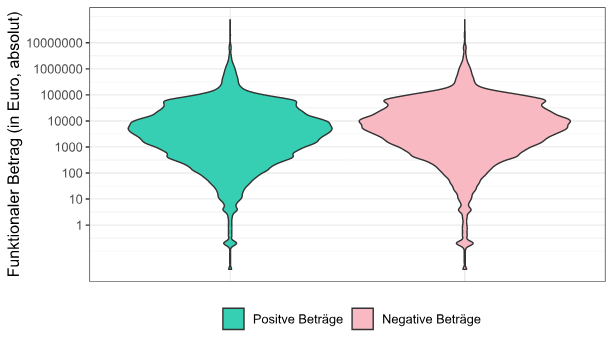


Abbildung 7. Verteilung der funktionalen Beträge nach positiven und negativen Beträgen.

…

## 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.

1. Im Rahmen dieser Ausarbeitung wird der Begriff „Buchung“ synonym mit „Buchungszeile“ verwendet. [↑](#footnote-ref-1)