

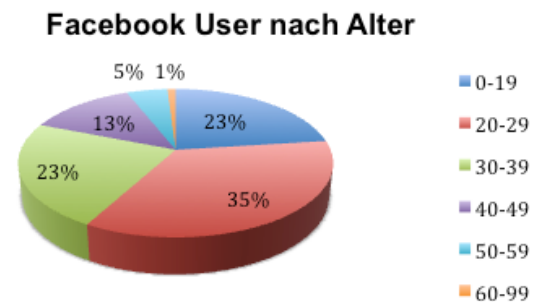
Modul 100: Daten charakterisieren, aufbereiten und auswerten

Daten analysieren und auswerten

Geschäftsgrafiken: Kreisdiagramm

Das Kreisdiagramm zeigt eine **ganze, abzählbare Menge**. Diese Menge wird in **Segmente** unterteilt. Die Segmente sind Teilmengen, wobei im Kreisdiagramm **Teilmengen** mit Kontrastfarben unterschieden werden. Die stärkste Kontrastfarbe bekommt das grösste Segment (rot), gefolgt von blau, grün, gelb usw. Es gilt, dunkle, markante Farben für grosse Segmente und helle Farben für die kleineren Segmente anzuwenden.

In folgendem Kreisdiagramm sehen Sie die Facebook User der Schweiz nach Altersgruppen unterteilt.



Quelle: Facebook Schweiz 2011

Folgende Schlussfolgerungen können aus dem Kreisdiagramm entnommen werden.

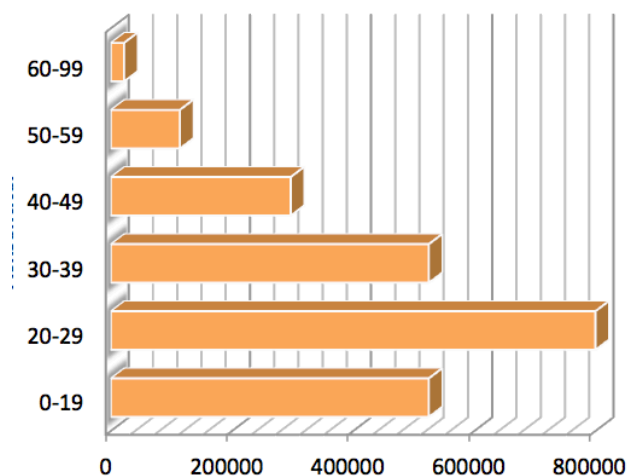
- Die grösste Altersgruppe von Schweizer Facebook (FB) Usern besteht aus 20 bis 29 Jährigen und macht demnach ein gutes Drittel aller User von FB aus.
- Teenager bis zum 19. Altersjahr sowie die 30 bis 39 Jährigen sind mit je einem Viertel die zweitgrösste Altersgruppe von FB in der Schweiz.
- Es gibt auf zehn Menschen mehr als einen, der zwischen 40 bis 49 ist und FB nutzt.
- Es gibt auf 100 Menschen gerade einmal fünf, die zwischen 50-59 sind und FB nutzen.
- Die über 60 jährigen Menschen nutzen praktisch kein FB.

Das Balkendiagramm

Das Balkendiagramm stellt zwei oder mehrere **Grössen im Vergleich** gegenüber. Es werden **absolute Werte** verglichen und nicht Verläufe.

Die Balken sind meistens von gleicher Breite und beschriftet.

Das Balkendiagramm zeigt die Anzahl Schweizer Facebook User pro Altersklasse in **absoluten Zahlen**.



Quelle: Facebook Schweiz 2011

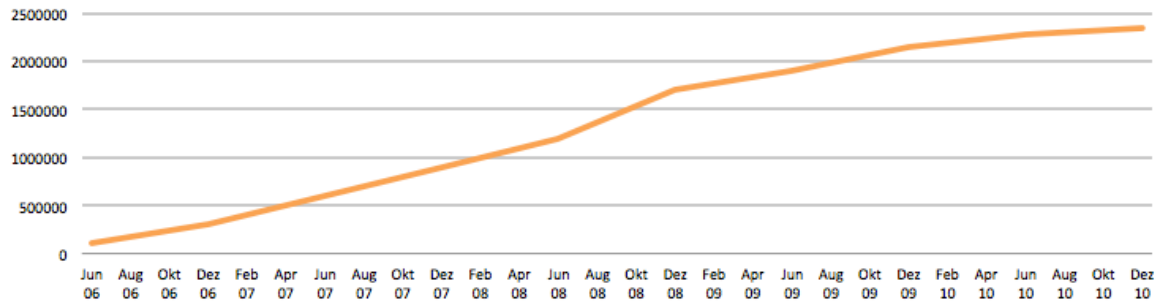
Man sieht sehr schön, dass 2'283'000 von 8 Millionen Schweizern Facebook nutzen. Das entspricht ca. 29% aller Schweizer.

Was können Sie dem Balkendiagramm noch entnehmen?

Das Liniendiagramm

Das Kurvendiagramm zeigt **Entwicklungsverläufe über eine bestimmte Zeitspanne**. Mit zwei Linien werden zu **vergleichende Entwicklungsprozesse** gegenübergestellt. Oft wird das **kartesische Koordinatensystem** eingesetzt. Die vertikale Achse nennt man **Ordinate**, die horizontale Achse **Abszisse**. Jede Achse wird mit einer eindeutigen Grösse mit Masseinheit beschriftet und die Skala passend unterteilt.

Folgendes Liniendiagramm zeigt den Zuwachs an Facebook Usern in der Schweiz.



Quelle: Facebook Schweiz 2011

Man sieht eine stetige Wachstumskurve, die seit 2009 an Steigung leicht einbüsst. Das heisst im Jahre 2010 gab es weniger Neuregistrierungen bei Facebook als im Jahre 2009.

Die Tabelle

Die Tabelle ist eine strukturierte Form eines Datenbestands. Man könnte sagen, jedes Element einer Tabelle hat einen Ort in der Tabelle, wo es hingehört. Wir unterscheiden später schwach strukturierte und stark strukturierte Tabellen.

Schauen wir uns die Tabelle und die dazugehörige Terminologie mit Hilfe eines Beispiels an:

Arbeitsbericht: Downey Ronny

Datum	Arbeitsbezeichnung	Zeitaufwand
05.06.2012	Büroarbeit (Korrespondenz, E-Mail)	40 min.
	Telefon mit Kunde	30 min.
07.06.2012	Telefon mit Chef	1 h
	Emails	40 min.
11.06.2012	Büroarbeit (Korrespondenz, E-Mail)	30 min.
	Telefon mit Kunde O'Brien	40 min.
	E-Mails	30 min.
20.06.2012	Chef geholfen	40 min.
25.06.2012	Telefone, Fragen geklärt	30 min.
28.06.2012	Büroarbeit (Korrespondenz, E-Mail)	45 min.
	Telefon mit Kunde O'Neely	40 min.

Die obige **Tabelle**, auch **Entitätsmenge** genannt, hat einen eindeutigen **Tabellennamen**. Sie heisst Arbeitsbericht: Downey Ronny.

Sie besteht aus den drei **Spalten** oder im Fachbegriff aus den drei **Attributen** Datum, Arbeitsbezeichnung und Zeitaufwand. Dabei ist jeder Attributsname (Spaltenname) eindeutig.

Das Attribut Datum **identifiziert** zudem jedes Tupel der Tabelle durch einen einmaligen Eintrag. Dieses Attribut heisst **Primärschlüssel** oder **Identifikationsschlüssel**.

Die Tabelle verfügt über sechs **Zeilen** (**Entität**, **Tupel**, **Datensatz**) mit Einträgen. Wir nennen diese **Zelleninhalte Datenwerte** oder **Values**.

Definition Tabelle

1. Eine Entitätsmenge (Tabelle) besteht aus einem eindeutigen Tabellennamen.
2. Innerhalb einer Tabelle ist jeder Attributname eindeutig und bezeichnet eine bestimmte Spalte.
3. Ein Attribut identifiziert eindeutig die Tupel in einer Tabelle und wird als Primärschlüssel bezeichnet.

Die Anzahl der Tupel einer Tabelle ist beliebig, die Anordnung der Tupel innerhalb einer Tabelle bedeutungslos.

Die genaue und vollständige Tabellendefinition wird im Modul 104 im zweiten Lehrjahr erarbeitet. Für Modul 100 reicht die vorliegende Definition aus.

Wir beschäftigen uns vorwiegend mit strukturierten Daten.

Doch was machen strukturierte Daten aus?

Worin unterscheiden sie sich von unstrukturierten Daten?

Strukturierte Daten

Felder, Datensätze, Dateien, Tabellen

Programme, Modelle, Skript

Stützen sich auf Tools, Datenbanken

Schwierig ohne Tools zu verwalten

Unstrukturierte Daten

Dokumente, Fotos, Grafiken, Text, Video, Chat

Schwierig Werkzeuge zu finden, obschon mehr und mehr davon auf den Markt gebracht werden.

Dokumentmanagement Tools

Können ohne Tools verwaltet werden.

Allgemein unterscheidet man zwischen strukturierten und unstrukturierten Daten. Strukturierte Daten beinhalten Informationen, welche in mehrfachen und strukturierten Formen gespeichert werden. Diese Art der Sammlung von Daten wird für Dateien, Tabellen, Datenbanken, Data Warehouses etc. verwendet. Sie lassen sich relativ einfach maschinell weiterverarbeiten.

Im Gegensatz dazu können Menschen mit unstrukturierten Daten leichter umgehen. Unstrukturierte Daten sind Dokumente, Bilder, Grafiken, Video, Audio etc. Sie lassen sich nur schwer bzw. ungenau maschinell weiterverarbeiten.

Repetitorium AB100-02

Grundkenntnisse: Datencharakterisieren

Begriffe:

- | | | | |
|----------|-----------------|----------|----------------|
| A | Primärschlüssel | B | Absoulte Werte |
| C | Datenwert | D | Liniendiagramm |
| E | Dokumente | F | Ordinate |
| G | Balkendiagramm | H | Attributnamen |
| I | Entitäten | J | Kreisdiagramm |

Welcher der oben aufgeführten Begriffe passt jeweils am besten zu den unten aufgeführten Beschreibungen? Mehrfachnennungen sind möglich (ein Begriff passt zu mehreren Beschreibungen).

Beschreibungen:

Die vertikale Achse des kartesischen Koordinatensystems heisst...

A B C D E F G H I J

Dass die Schweiz bald die 8 Millionen Einwohnergrenze erreicht ist ausgedrückt in...

A B C D E F G H I J

Alle Zeilen einer Tabelle werden durch diesen eindeutig identifiziert...

A B C D E F G H I J

Neben dem Tabellennamen werden auch die Spalten eindeutig benannt.

A B C D E F G H I J

Stellt Segmente in Vergleich zu einer Gesamtmenge.

A B C D E F G H I J

Der Zelleninhalt einer Tabelle heisst...

A B C D E F G H I J

Eine Tabelle ist eine Menge von ...

A B C D E F G H I J

Braucht neben dem Liniendiagramm auch das kartesische Koordinatensystem.

A B C D E F G H I J

Den täglichen Kursverlauf einer Aktie verfolgt man am besten mit einem ...

A B C D E F G H I J

... sind unstrukturierte Daten.

A B C D E F G H I J

Auftrag

Der Arbeitsrapport von Ronny Downey ist in einer schlecht strukturierten Tabelle vorgestellt worden. Zwar liessen sich wichtige Elemente einer Tabellen definieren und einführen, aber aus Sicht der relationalen Datenbankmodellierung müsste einiges verbessert werden.

Aus Sicht eines Datenbankspezialisten fällt sofort auf, dass die Tabelle über **nicht atomare Datenwerte** (Zelleninhalte) verfügt.

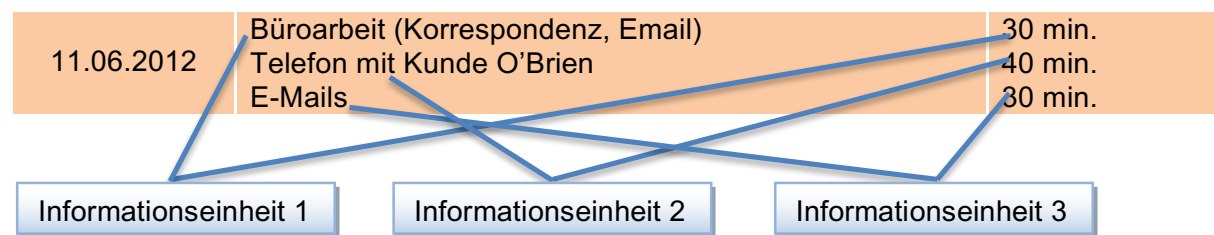
Definition

Atomare Datenwerte sind Zelleninhalte, die sich in keine weiteren Informationseinheiten unterteilen lassen.

Betrachten wir Zeile 3:

Die zweite und dritte Zelle bestehen aus jeweils drei Informationseinheiten:

2. Zelle	3. Zelle
<ul style="list-style-type: none"> • Büroarbeit (Korrespondenz, E-Mail) • Telefon mit Kunde O'Brien • E-Mails 	<ul style="list-style-type: none"> • 30 min. • 40 min. • 30 min.



In der relationalen Datenbankmodellierung ist das Vermeiden von nicht atomaren Datenwerten eine wichtige Entwurfsregel.

Nicht atomare Datenwerte müssen beim Tabellenentwurf beseitigt werden.

Der Chefin von Ronny Downey reichen zudem die Informationen aus dem Arbeitsrapport nicht aus, um die Leistungen fair zu bewerten. Sie verlangt folgende Tabellenerweiterungen:

1. Die Zeilen sollen nicht durch das Datum sondern durch eine numerische Aufzählung pro Datensatz identifiziert werden.
2. Die Unterteilung der Arbeit in Tätigkeiten (Büroarbeit, Telefon, Email, ...) ist gut. Die Chefin wünscht, dass jede Tätigkeit zusätzlich mit einem Aufgabenziel beschrieben wird. Beispiel: Telefon mit O'Donnell's: Ziel – Terminvereinbarung für Vertrag
3. Neben dem Aufgabenziel soll ein Arbeitsstatus erfasst werden. Ein Status ist ein Kurzbeschreibung mit dem momentanen Stand der Arbeiten.

- Erarbeiten Sie eine Dokumentvorlage für einen entsprechenden Arbeitsrapport in Microsoft Excel oder OpenOffice Calc aus.

- Erfassen Sie weitere Daten zur Veranschaulichung Ihrer Vorlage.

- Senden Sie Ihre Lösung der Lehrperson zur Kontrolle.

Zusatzaufgabe:

4. Die Chefin will schnell ein Kreisdiagramm erstellen können, indem zu sehen ist, wie viel Arbeitszeit Ronny Downey für Büroarbeiten oder E-Mails u.a. benötigt hat.
5. Die aufgewendete Zeit pro Arbeitstätigkeit soll in absoluten Zahlen in einem Balkendiagramm dargestellt werden können.