

## Data PreProcessing:

- Zusammenfügen der Datensätze & Sicherstellung, dass richtig angefügt
- Missing-Behandlung
- Ausreißer-Behandlung

→ *Siehe Jupyter Notebook*

## Planung der Datenanalyse:

Wir haben einen starken Fokus auf die Frage gelegt, von was ein erfolgreiches Funding abhängt.

### Neue Spalten in Vorbereitung auf die Datenanalyse:

Hierzu wurde der bestehende Datensatz um mehrere Spalten erweitert: Einerseits wurden die Funding-Ziele und die Funding-Höhe jeweils kategorisiert in 10 Kategorien, um sich Häufigkeiten der verschiedenen Höhen besser grafisch abbilden zu können. Des Weiteren wurden zwei zusätzliche Spalten `successful_funding` sowie `success_factor` erstellt. `Successful_Funding` besitzt 4 Kategorien:

1. `No_funding`: Es kam überhaupt kein Funding zustande.
2. `Not_achieved`: Die volle Höhe des Funding-Ziels wurde nicht erreicht.
3. `Achieved`: Die Höhe des Funding-Ziels wurde genau erreicht.
4. `Goal_overachieved`: Das Funding-Ziel wurde sogar übertroffen, d.h. es wurde mehr finanziert als von den Initiatoren als Funding-Ziel angegeben wurde.

Die Spalte `success_factor` misst den Anteil des geleisteten Fundings am Funding-Ziel. Bei der Visualisierung hat sich gezeigt, dass die Rechenleistung für bestimmte Plots (z.B. `swarm-Plot`) in Verbindung mit dem `success_factor` (z.B. mit `gender_initiators`) nicht ausreichte. Auch ein neues Dataframe mit den nur relevanten Spalten konnte das Problem nicht beheben. Deshalb wurde vorwiegend mit den vier Kategorien von `successful_funding` gearbeitet.

Die Spalte `use` war aufgrund ihrer sehr unstrukturierten Daten für quantitative Auswertungen unbrauchbar, jedoch sollte untersucht werden, ob die Angabe bzw. das Weglassen eines „`use`“ beim Crowdfunding bzgl. der Funding-Höhe und des Funding-Erfolgs Vor- bzw. Nachteile bringt. Daher wurde die Spalte `use_description` erstellt, welche das Vorliegen einer `use` bzw. das Fehlen einer `use-description` abbildet. Schnell wurde festgestellt, dass nur in sehr wenigen Fällen keine `use-description` angegeben wurde und diese Spalte damit über wenig Aussagekraft verfügt.

Die Spalte `borrower_genders` war aufgrund ihrer Zusammensetzung unbrauchbar, jedoch war ersichtlich, dass sogar 2 wertvolle Informationen in der Spalte enthalten sind:

1. Die Anzahl der Initiatoren
2. Die Geschlechterverteilung der Initiatoren

Um mit diesen Informationen arbeiten zu können, wurden die Spalten `gender_initiators` und `count_initiators` erstellt. Als Zwischenschritt mussten die Spalten `count_female` und `count_male` erstellt werden. Leider waren viele Visualisierungen, die Zusammenhänge zwischen der Initiatoren-Anzahl bzw. Initiatoren-Geschlechterverteilung und Erfolg des Fundings abbilden sollten, mangels Rechenleistung nicht umsetzbar. Daher musste auf einfachere Visualisierungen zurückgegriffen werden.

### **Leistungsoptimierung durch Anpassung der Datentypen:**

Auch die Anpassung der Datentypen (von float bzw. int64 auf int32) hat eine sehr marginale Beschleunigung gebracht.

### **Analyse verschiedener Fragestellungen (Visualisierung):**

Folgende Fragen leiteten die Analyse:

1. Gibt es Auffälligkeiten bei der Anzahl der Projekte in Abhängigkeit von verschiedenen Variablen?
2. Wovon hängen der Erfolg und die Höhe des Fundings ab? Fokussiert wurde dabei auf die Frage, welchen Einfluss die Größe und Geschlechterzusammensetzung des Initiatoren-Teams auf das Funding bzw. den Erfolg des Fundings hatte.
3. Zeigen sich in der Teamzusammensetzung der Initiatoren-Teams Muster?

Auf der nachfolgenden Seite finden sich die jeweiligen Ergebnis-Grafen, sowie die Aussagen, die anhand der Grafen getroffen werden können.

# Ergebnisse

Die Befehle sind dem Jupyter-Notebook zu entnehmen.

## Frage 1: Gibt es Auffälligkeiten bei der Anzahl der Projekte in Abhängigkeit von verschiedenen Variablen?

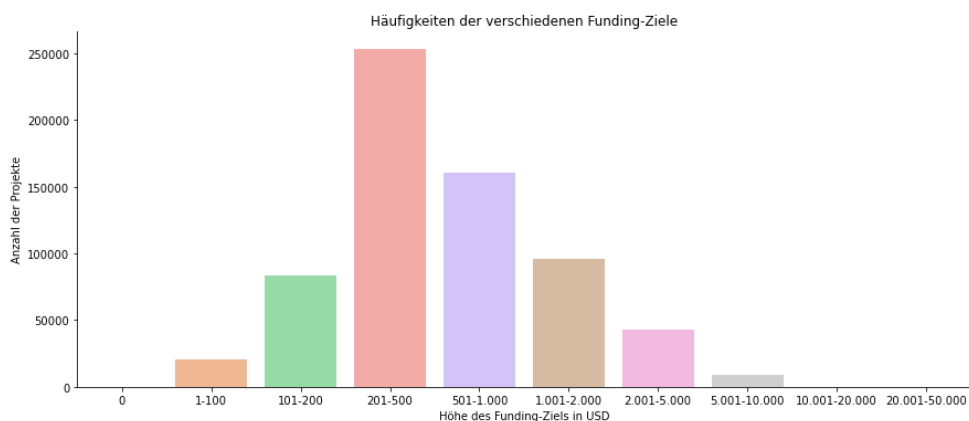
Anzahl der Projekte in Abhängigkeit von der/dem:

### a) Erfolg des Fundings

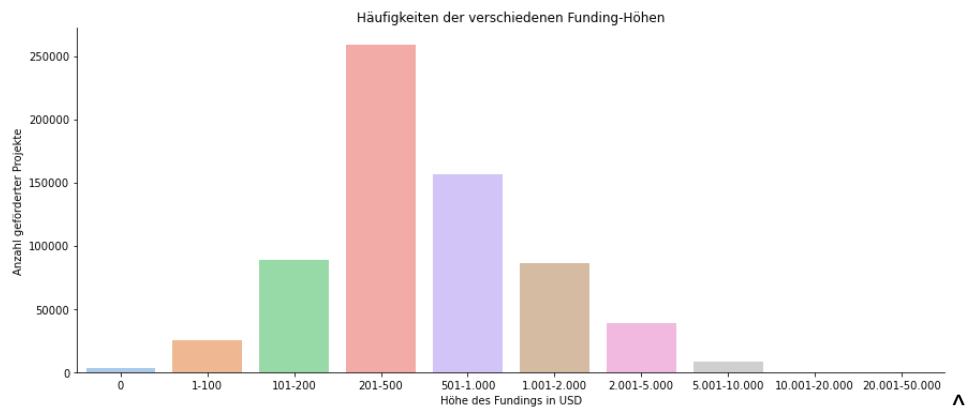


Das Funding-Ziel wurde bei den allermeisten Projekten (über 600.000 Projekte) erreicht. Nur in wenigen Fällen (< 5.000 Fälle) wurde das Funding-Ziel nicht erreicht, d.h. es wurde weniger Geld gefundet als von den Initiatoren als Ziel vorgegeben wurde. Bei einem nur sehr kleinen Prozentsatz der Projekte kam überhaupt kein Funding zustande. In zwei Fällen der knapp 700.000 Projekte wurde das Funding-Ziel sogar übertroffen (mehr Geld eingesammelt als notwendig). In dieser Grafik sind diese kaum ersichtlich, da sie einen verschwindend kleinen Anteil ( $2.85E-6$ ) an der Gesamtzahl der Projekte ausmachen.

### b) Funding-Zielen



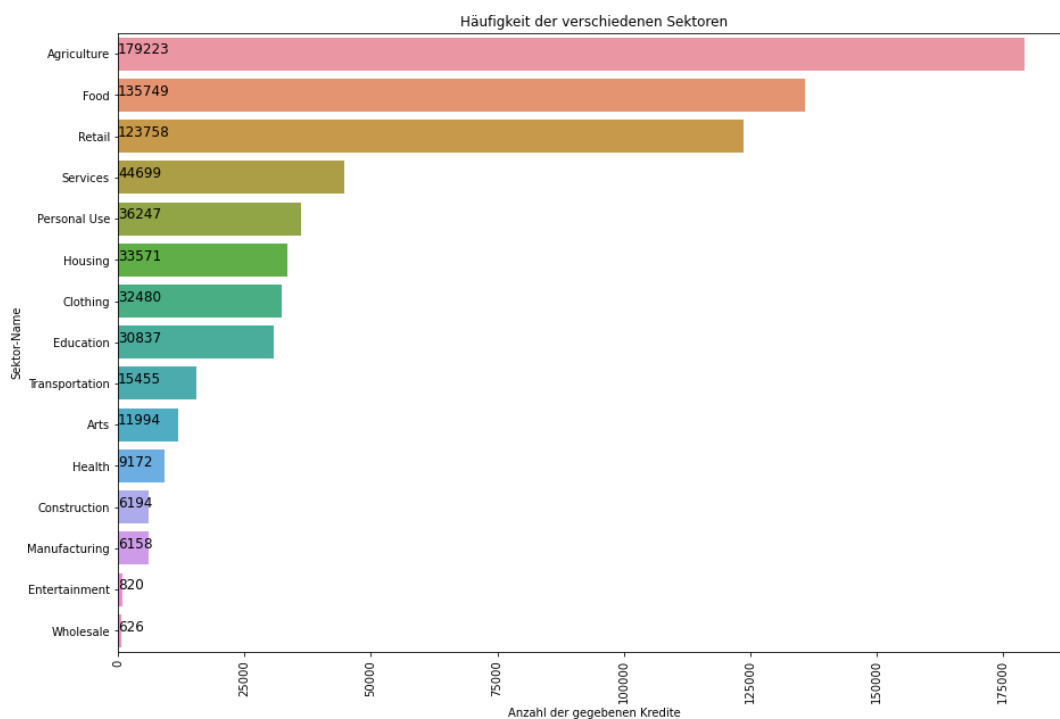
### c) Funding-Höhen



Die Grafiken zeigen, dass die meisten Projekte mit einer Höhe von 201-500 USD angesetzt (Grafik oben) und dann auch finanziert wurden (Grafik unten). Projekte über 5.000 USD sind eher die Seltenheit. Crowd-Funding scheint eher für Projekte mit "kleineren" Beträgen interessant zu sein.

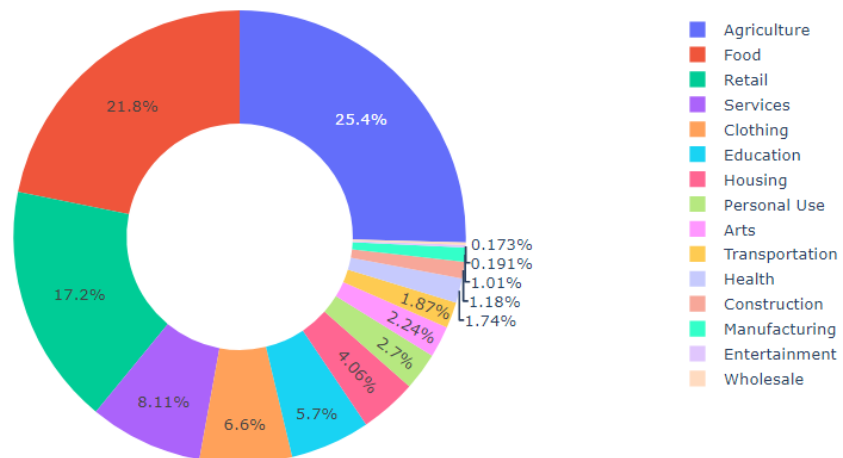
Ein Unterschied zwischen beiden Grafiken ist fast nicht erkennbar. Abzulesen in Grafik 2 ist jedoch, dass einige Projekte, die mit einem Betrag über 0 USD angesetzt wurden, offensichtlich kein Funding erhielten.

#### d) Sektoren



In den Sektoren Agriculture (179.223 Projekte), sowie Food (135.749 Projekte) und Retail (123.758 Projekte) werden mit Abstand die meisten Crowdfunding-Projekte realisiert.

Total Funding Amount in percent per Sector



```
1 sum(funding_per_sector)
523567710
```

Insgesamt wurden 523 Mio. USD Gelder in verschiedenste Projekte investiert.

Davon wurde mit knapp 133 Mio. USD (siehe Hover-Effekt) bzw. 25,4% der insgesamt investierten Gelder am meisten Geld in den Sektor Agriculture investiert. Gefolgt wird der Sektor Agriculture von den Sektoren Food mit 114 Mio. USD (21,8% aller Fundings in USD) und Retail mit 90 Mio. USD (17,2%). Die übrigen Sektoren machen weniger als 50% an der Gesamtfunding-Summe aus.

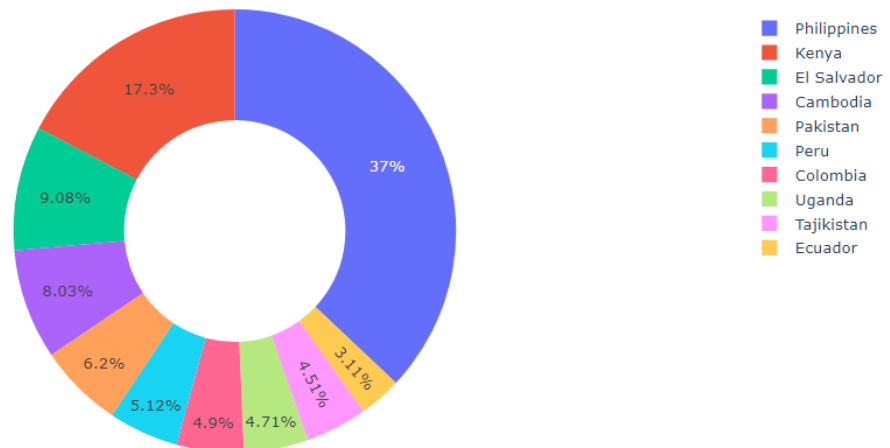
## e) Ländern

### Anzahl der Projekte in verschiedenen Ländern



Die Größe und Häufigkeit des Landes Philippinen wie auch El Salvador und Kenya zeigen, dass hier viele Projekte finanziert werden.

## Anzahl an Projekten der Top 10 Länder

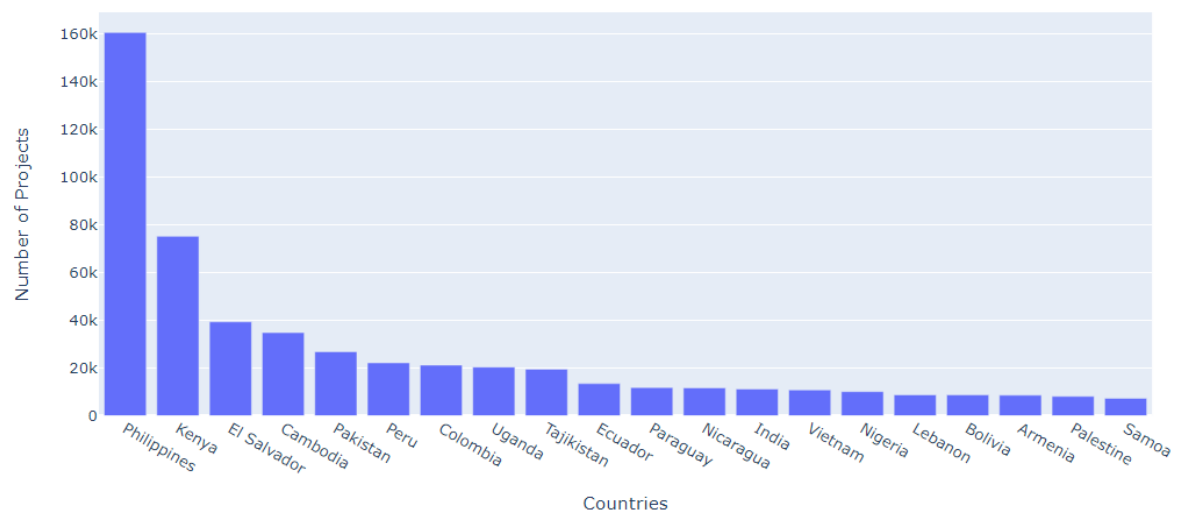


In den Philippinen und Kenya finden mit 37% aller Projekte (160.361 Stück) respektive 17,3% (75.113 Stück) mit Abstand die meisten Crowdfunding-Projekte statt!

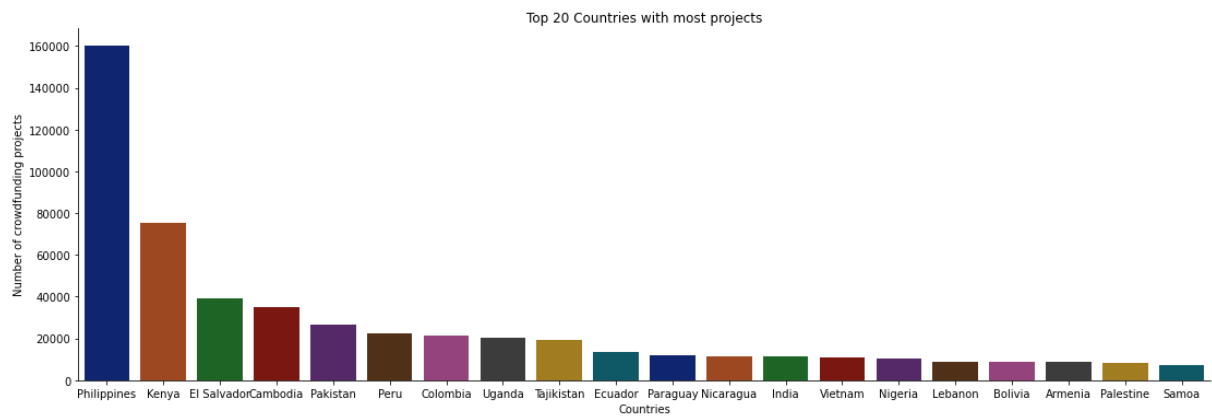
Alternativer Plot, falls man etwas mehr Länder anzeigen will:

Mit plotly express:

Top 20 Countries with most projects

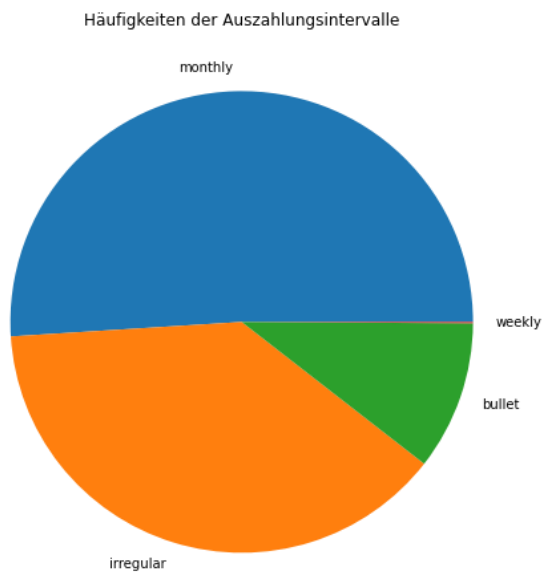


Mit Matplotlib:



Auch die Balkendiagramme zeigen, dass in den Philippinen und Kenya mit Abstand die meisten Crowdfunding-Projekte stattfinden. Scheinbar wird Crowdfunding hauptsächlich von Schwellen- und Entwicklungsländern genutzt. Das spielt sich auch in der Häufung von niedrig Funding-Zielen (am meisten zwischen 200-500 USD) wieder.

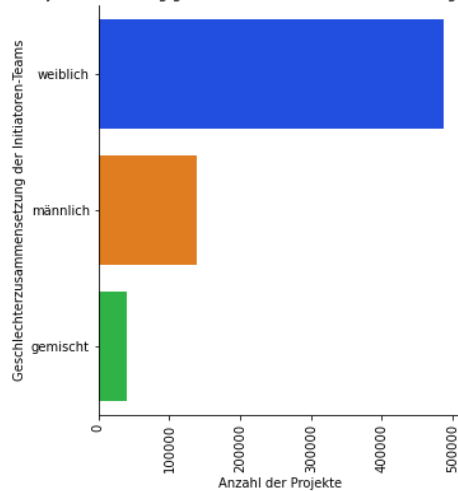
## f) Auszahlungsintervallen



Die Fundings werden ca. zur Hälfte in monatlichen Rückzahlungsintervallen zurückgezahlt. Ein unregelmäßiges Rückzahlungsintervall ist mit ca. 30% am zweithäufigsten vertreten. Am seltensten ist die wöchentliche Rückzahlung. Rückzahlungen als Einmalzahlung sind mit ca. 10% vertreten.

## g) Geschlechterzusammensetzung der Initiatoren-Teams

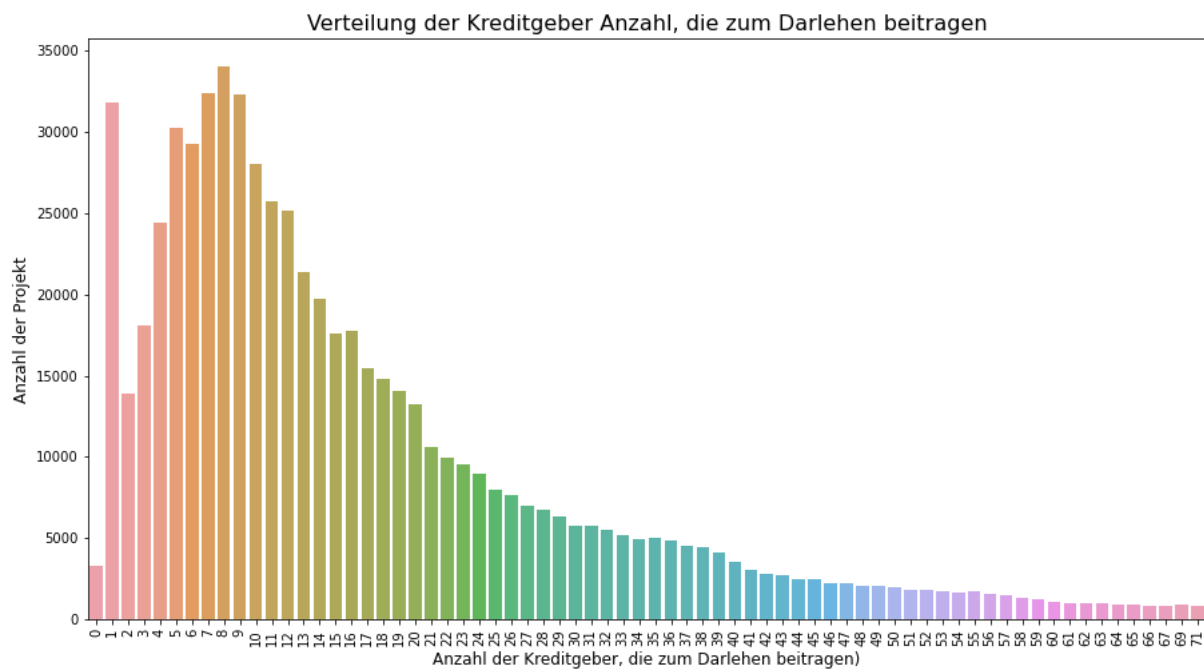
Anzahl der Projekte in Abhängigkeit von der Geschlechterverteilung der Initiatoren-Teams



Der Plot zeigt, dass die meisten Projekte von reinen Frauenteam geleitet werden.



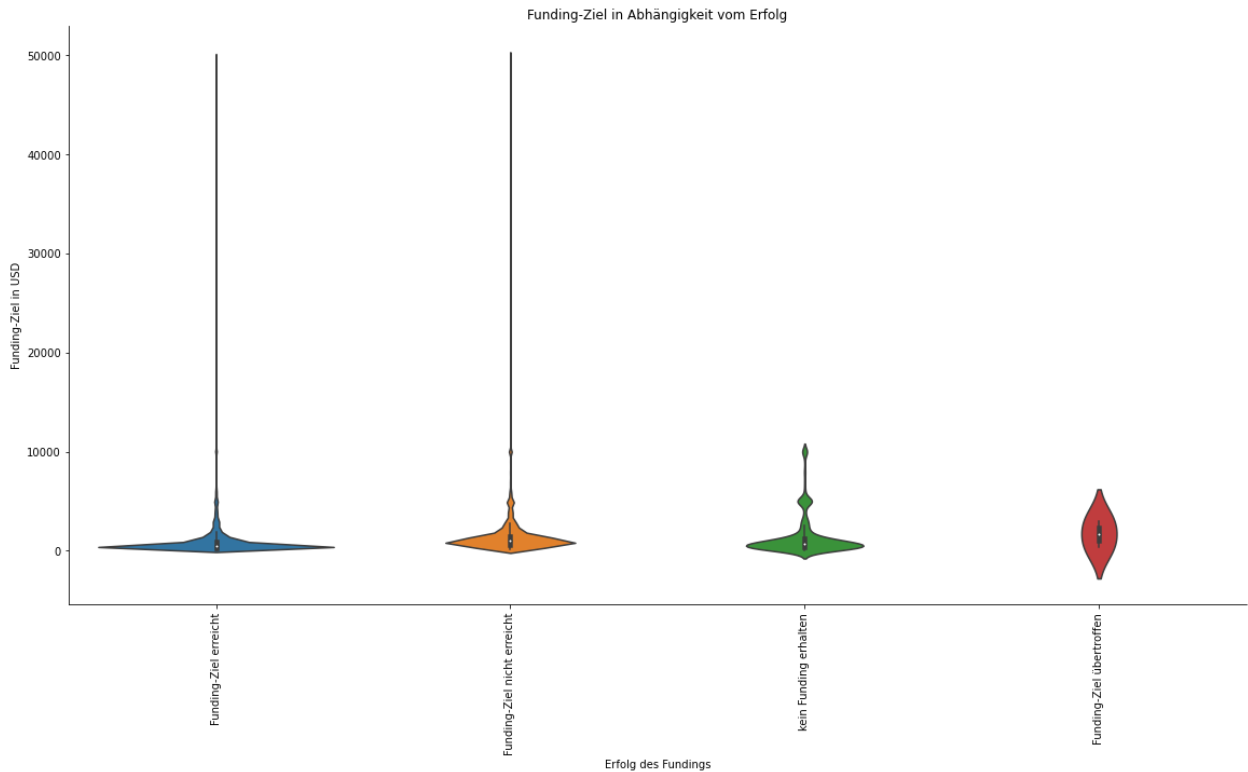
## h) Anzahl der Kreditgeber



Die Verteilung ist stark verzerrt: Die meisten Projekte haben 8 Kreditgeber. Auffällig ist, dass auch Projekte mit einem Kreditgeber besonders häufig sind.

**Frage 2: Wovon hängen der Erfolg und die Höhe des Fundings ab?** Fokussiert wurde dabei auf die Frage, welchen Einfluss die Größe und Geschlechterzusammensetzung des Initiatoren-Teams auf das Funding bzw. den Erfolg des Fundings hatte.

### 1. Sind teurere oder billigere Projekte erfolgreicher?

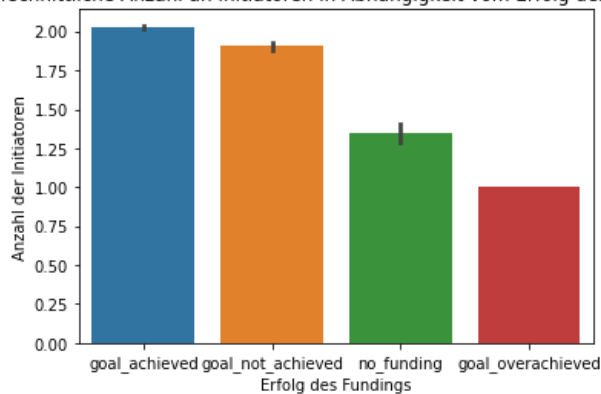


Die Grafik zeigt, dass es klare Häufungen bei Projekten mit niedrigerem Funding-Ziel gibt. Bei besonders hohem Funding-Ziel wurde immer ein Funding gegeben: entweder in voller Höhe (= Funding-Ziel erreicht) oder ein geringerer Betrag finanziert (Funding-Ziel nicht erreicht).

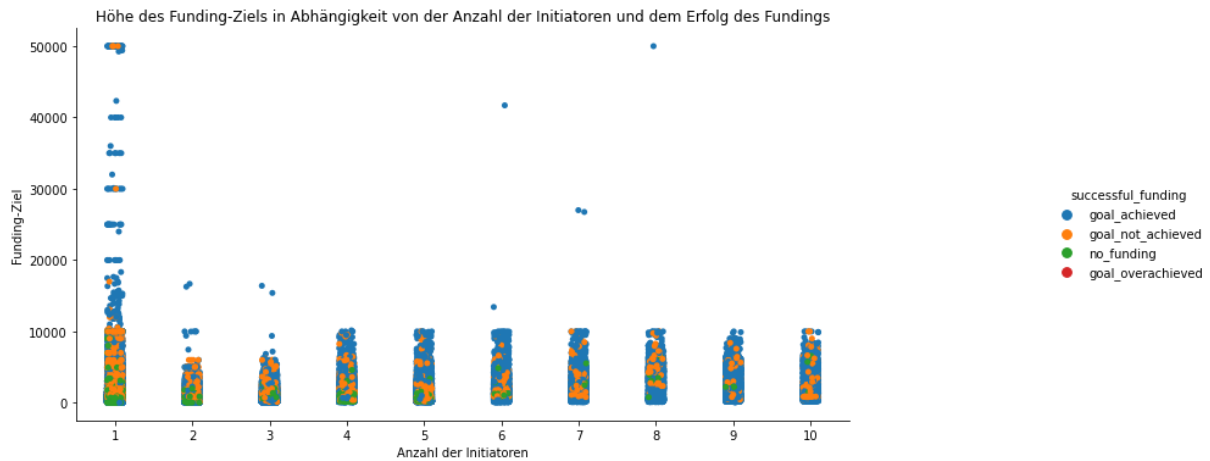
### 2. Welchen Einfluss haben die Eigenschaften des Initiatoren-Teams auf das Funding?

#### a) Welchen Einfluss hat die Anzahl der Initiatoren?

Durchschnittliche Anzahl an Initiatoren in Abhängigkeit vom Erfolg des Fundings

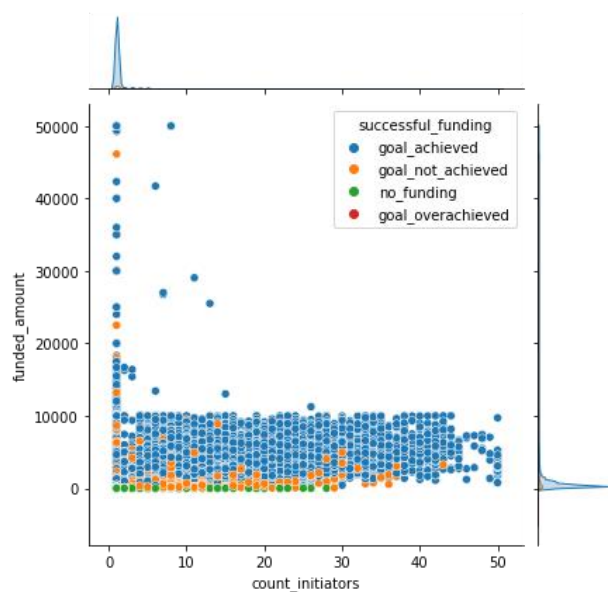
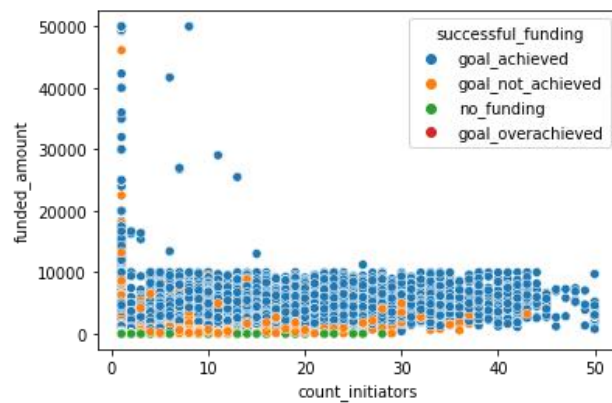


Die Projekte, die ihr Funding-Ziel erreichten hatten im Schnitt am meisten Initiatoren. Die Projekte, die kein Funding erhielten, hatten im Durchschnitt weniger Initiatoren. Projekte, die sogar mehr Geld erhielten, als sie brauchten wurden immer von einer Person initiiert (mean=1)!



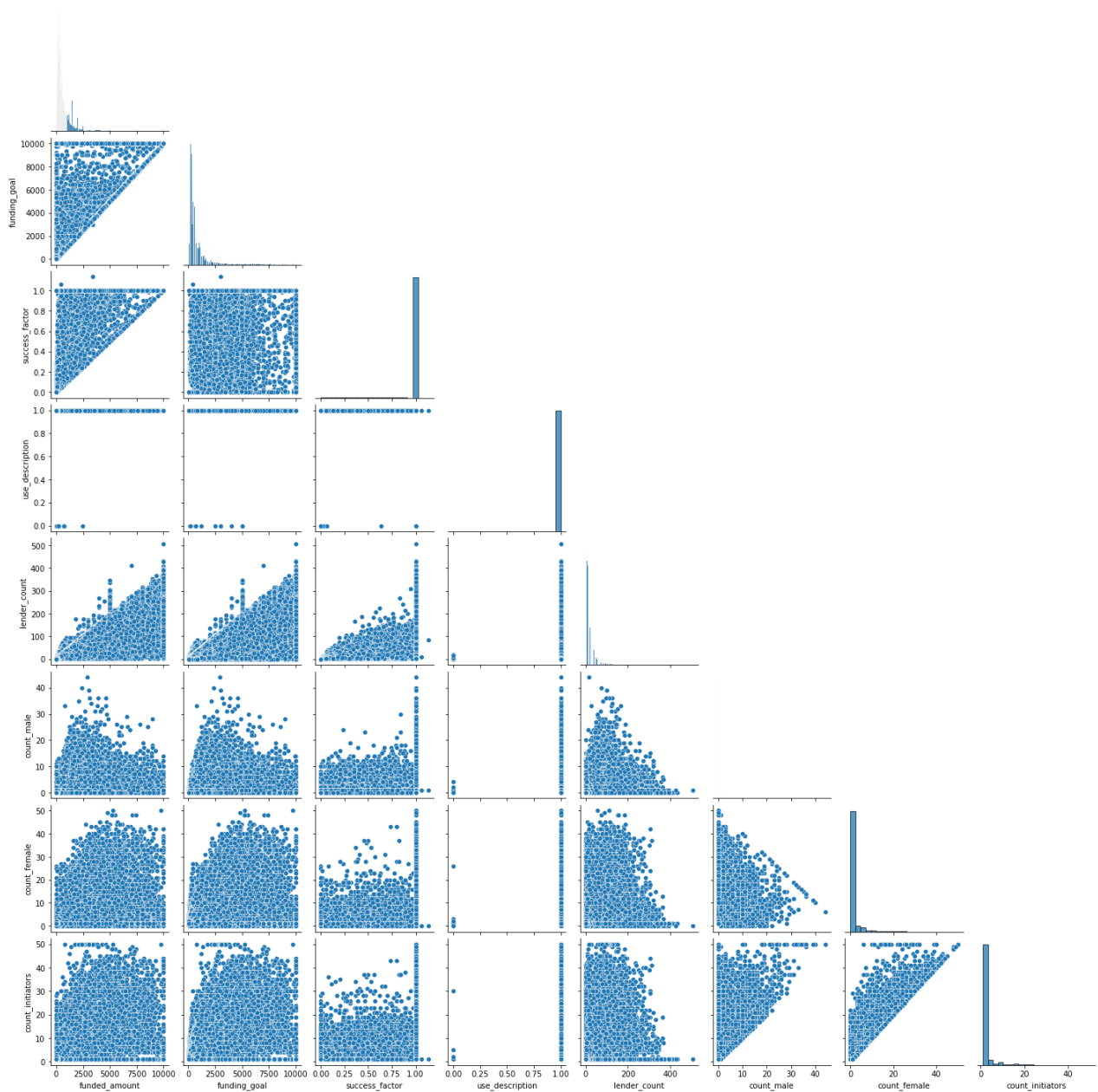
Hier wurden nur die Gruppengrößen mit 10 Initiatoren betrachtet, da eine vorherige Betrachtung aller Gruppengrößen zeigte, dass die größten Auffälligkeiten in diesem Bereich zu finden sind. Das Funding-Ziel scheint am häufigsten bei Gruppengrößen zwischen 5 und 7 erreicht worden zu sein.

### Gezahltes Funding in Abhängigkeit der Anzahl der Initiatoren und Erfolg des Fundings



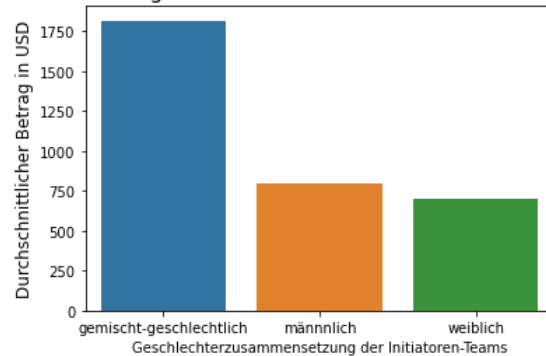
Obwohl es sehr wenige Projekte über 10.000 USD gibt, häufen sich diese bei wenigen Initiatoren. Die meisten Projekte liegen unter 10.000 USD. Bei sehr vielen Initiatoren (mehr als 30) kam immer ein Funding zustande, scheinbar jedoch nur im Bereich bis 10.000 USD.

Es wäre in einer weiteren Analyse sicherlich noch sinnvoll, sich nur die Projekte unterhalb von 10.000 USD anzusehen, da wir es hier mit besonders vielen Projekten zu tun haben. Einen ersten Überblick bietet dieser Pairplot, auf den wegen weiterer Fragestellungen jedoch nicht weiter eingegangen wurde:



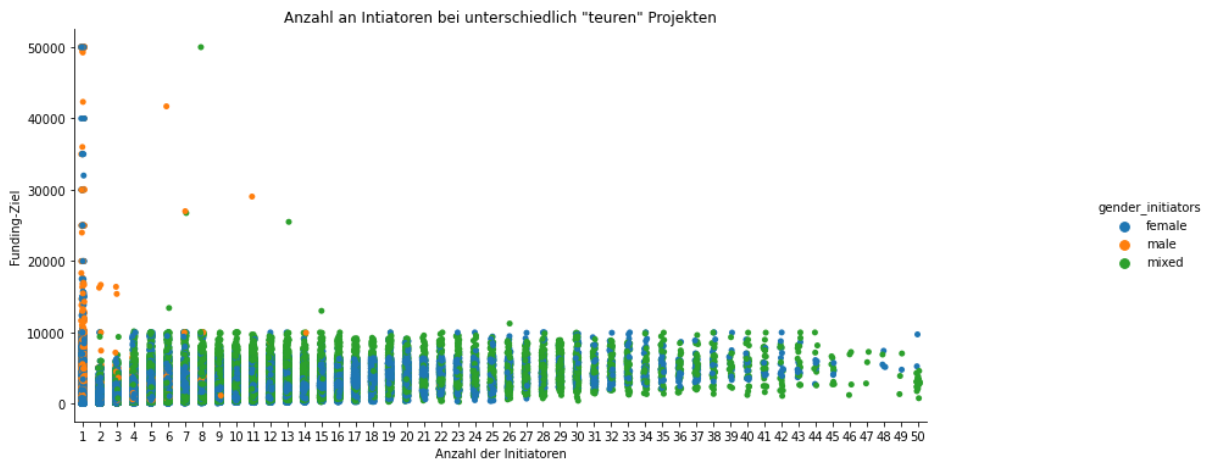
## b) Welchen Einfluss hat die Geschlechter-Zusammensetzung des Initiatoren-Teams?

Durchschnittliche Finanzierung nach Geschlechterzusammensetzung der Initiatoren-Teams



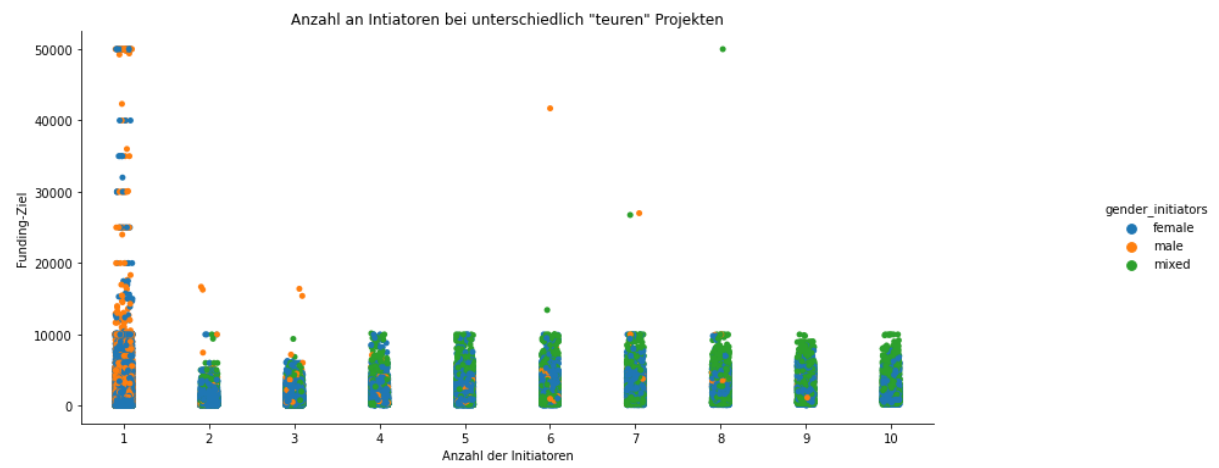
Die Grafik zeigt, dass gemischt-geschlechtliche Teams im Durchschnitt höhere Finanzierungen erhalten als rein männliche oder rein weibliche Teams. Weibliche Teams fahren dabei im Durchschnitt die geringsten Finanzierungsbeträge ein.

## c) Welchen Einfluss haben Anzahl der Initiatoren in Kombination mit der Geschlechter-Zusammensetzung?



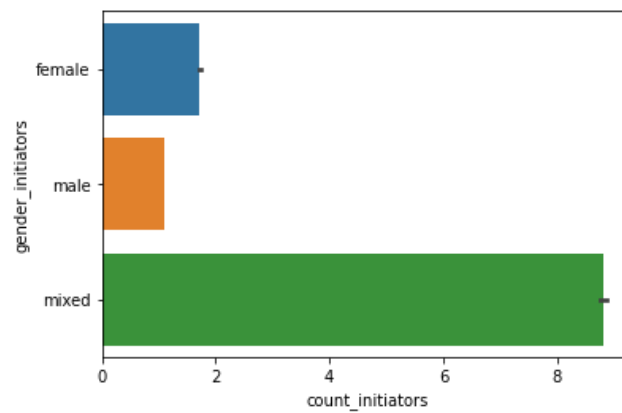
Der Plot zeigt, dass Projekte, die über 10.000 USD benötigen, vorwiegend von einer Person geleitet wurden. Interessant ist, dass Männer sich das scheinbar eher zutrauen: unter den alleinigen Projektleitern und denen, die hohe Beträge benötigen häufigen sich die Männer! Frauen und gemischte Teams leiten dafür viele Projekte. Das entspricht auch psychologischen Erkenntnissen, dass sich Männer im Schnitt mehr zutrauen als Frauen und Frauen im Schnitt bessere Team-Player sind.

Diese Vergrößerung des Plots zeigt das sehr deutlich:



### Frage 3: Zeigen sich in der Teamzusammensetzung der Initiatoren-Teams Muster?

#### a) Zusammenhang zwischen Größe des Teams und der Geschlechterzusammensetzung



Rein weibliche Teams bestehen maximal aus 2 Frauen. Davon scheint es aber sehr viele zu geben (siehe Grafik darüber!). Männer arbeiten entweder alleine oder in gemischt-geschlechtlichen Teams, jedoch nie in reinen Männer-Teams! Teams mit mehr als 2 Mitgliedern sind stets gemischt-geschlechtlich.