

# 1 Names

Hester Verdenius, Xavier van Dommelen, Floris Kappeyne.

## 2 Research questions/Hypothesis

### 2.1 Hoofdvraag

*Wat voor genre heeft en van welke publisher komt de "big hit" video game in 2017 voor elke van de volgende: Xbox, Playstation, Nintendo en PC?*

We kunnen deze voorspellingen dan vergelijken met de realiteit.

### 2.2 Definities

**Big hit:** Een video game met de hoogste sales in zijn categorie.

**Over tijd:** Per game release jaar.

### 2.3 Onderzoeks vragen

1. Hoe analyseren we onze data met betrekking tot de hoofdvraag?
2. Is er correlatie tussen:
  - Platform en Publisher
  - Platform en Sales
  - Platform en Genre
  - Platform en Score
  - Publisher en Sales
  - Publisher en Genre
  - Publisher en Score
  - Sales en Genre
  - Sales en Score
  - Genre en Score
3. Is er verschil in sales tussen continenten?
4. Wat zijn de "Big hits" per jaar?

### 3 Dataset and some characteristics

De sales data-set bevat informatie over de sales van video games uitgebracht vanaf 1980 t/m 2016:

Rank, Sales rank, Name of game, Platform for which it was released, Year of release, Game genre, Company which released the game, NA Sales, North American sales, EU Sales, European sales, Japan sales, Sales in other countries, Sales around the worlds

De reviews data-set bevat informatie over de IGN reviews van video games uitgebracht tussen 1997 en 2017:

Score (as phrase and number between 1-10), Name, Genre, Platform, Date of release, Editors Choice, url to the review on IGN

### 4 Time planning

Milestones per week:

- Week 1:
  - Eigen data set creëren uit de initiële datasets.
  - Aantal grafiekjes maken die een overzicht geven van de data om gevoel te krijgen over wat voor data we hebben.
  - "Big Hits" vinden uit 2017 om onze voorspellingen mee te vergelijken.
- Week 2:
  - Beginnen met het analyseren van de correlaties: goede technieken vinden, technieken uitvoeren.
  - Minimaal alle correlaties per Platform gevonden.
- Week 3:
  - De rest van de correlaties af voor **dinsdag**.
  - Techniek voor het voorspellen van "Big Hits" d.m.v. gevonden correlaties.
  - Begin van de poster (medium, layout, style).
- Week 4:
  - Clean up van het project.
  - Finaliseren poster.
  - Set up en oefenen presentatie.

Indien we tijd over hebben kunnen we kijken naar het verschil per continent, maar dan kunnen we niks meer doen met de reviews want die zijn wereldwijd.

## 5 Sources

Dataset sources:

1. Sales data: <https://www.kaggle.com/gregorut/videogamesales>
2. Review data: <https://www.kaggle.com/egrinstein/20-years-of-games>