

# Metody analýzy dat

## Analytická část projektu

ONDŘEJ ŘEHÁČEK (REH0063)

December 9, 2016

### 1 Dataset

Dataset obsahuje statistické informace o prodeji více než 16 tisíců videoherních titulů v rozdílných regionech. Je volně dostupný na internetové kegggle.com a byl publikován Gregory Smithem 10.Října 2016. Obsahuje tedy poměrně aktuální data čerpána z ověřeného zdroje ve videoherním světě - vgchartz.com, pomocí Python skriptu <sup>1</sup>. pro sběr dat z této webové stránky

#### 1.1 Motivace

Videoherní průmysl je již stejně dominantní jako filmový či hudební. Postupem času s vývojem techniky se vyvíjí i tato forma zábavy a v dnešní době mohou videohry vyprávět stejně komplexní příběhy jako celovečerní filmy. Není žádným tajemstvím, že jednotlivé kontinenty mají podstatně jinak rozložený prodej a zájem o videohry. S pomocí tohoto datasetu a jeho analýzy chci tyto rozdíly zdokumentovat a detailně popsat.

#### 1.2 Atributy

<b>Rank</b>	Pořadí titulu v počtu prodejů.
<b>Name</b>	Název herního titulu
<b>Platform</b>	Platforma na které titul běží.
<b>Year</b>	Rok vydání hry.
<b>Genre</b>	Žánr herního titulu.
<b>Publisher</b>	Vydavatel / Vývojář titulu.
<b>NA_Sales</b>	Prodeje na území státu serevní ameriky. (V milionech)
<b>JP_Sales</b>	Prodeje na území Japonska. (V milionech)
<b>EU_Sales</b>	Prodej na území státu Evropy. (V milionech)
<b>Other_Sales</b>	Prodej na území zbylých států. (V milionech)
<b>Global_Sales</b>	Celkový počet prodaných kusů.

<sup>1</sup><https://github.com/GregorUT/vgchartzScrape>

### 1.3 Úpravy a chybějící data

Přibližně 27 titulů bylo opraveno, protože obsahovaly nesmyslné hodnoty. Původně je v datasetu asi 329 položek, které nemají kompletní data - jedná se převážně o chybějící rok vydání. Některé z položek byly manuálně doplněny o zjištěná data.

### 1.4 Otázky, pro které hledáme odpovědi.

Kromě obecné analýzy dat, hledáme odpovědi na následující specifitější otázky:

- Jak se zvětšoval videoherní trh s postupem času ?
- Jaké jsou oblíbené žánry titulů v rozdílných regionech ?
- Jaké platformy byly nejspěšnější ve svých generacích ?
- Které žánry byly nejdominantnější na určitých platformách ?
- Které žánry jsou nejdominantnější v určitých regionech ?
- Preferují určité regiony rozdílné vývojáře ?

### 1.5 Analýza a Implementace

Analýza datasetu je provedena v R (IDE RStudio 1.0.44) a Python. Veškeré zdrojové kódy jsou přiloženy a byly využity následující balíčky:

- dplyr
- ggplot2

## 2 Obecná analýza

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

### 2.1 Počet vydaných her na rok.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis risus ante, auctor et pulvinar non, posuere ac lacus. Praesent egestas nisi id metus rhoncus ac lobortis sem hendrerit. Etiam et sapien eget lectus interdum posuere sit amet ac urna. Aliquam pellentesque imperdiet erat, eget consectetur felis malesuada quis. Pellentesque sollicitudin, odio sed dapibus eleifend, magna sem luctus turpis, id aliquam felis dolor eu diam. Etiam ullamcorper, nunc a accumsan adipiscing, turpis odio bibendum erat, id convallis magna eros nec metus. Sed vel ligula justo, sit amet vestibulum dolor. Sed vitae augue sit amet magna ullamcorper suscipit. Quisque dictum ipsum a sapien egestas facilisis.

### 2.2 Lorem ipsum dolor sit amet, consectetur adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis risus ante, auctor et pulvinar non, posuere ac lacus. Praesent egestas nisi id metus rhoncus ac lobortis sem hendrerit. Etiam et sapien eget lectus interdum posuere sit amet ac urna. Aliquam pellentesque imperdiet erat, eget consectetur felis malesuada quis. Pellentesque sollicitudin, odio sed dapibus eleifend, magna sem luctus turpis, id aliquam felis dolor eu diam.