Data at Invincibles Studio

Invincibles Studio

Max Lowe, Henry Childs
01/12/2025

Presentation Outline

- Jargon busting
- Invincibles Studio history
- The modern role of data in the gaming sector (and beyond)
- Real world examples of applied data in games
    - Analytics
    - Machine/Deep Learning
    - Engineering
- Industry overview for graduates

Invincibles Studio

Business Jargon Busting

- **UX** - User Experience - how a user interacts with your product. This includes design, usability and satisfaction
- **Events** - logged actions inside app (e.g. clicking a button or viewing a screen)
- **Database** - structured collection of data
- **Data Warehouse** - centralised large store of structured data
- **Data Lake** - unstructured raw data stores, typically extremely large
- **Big Data** - large data sets, requiring specialised tools to store, query and analyse
- **Agile** - current standard development methodology
- **CI/CD** - continuous integration & deployment pipelines, automation

Invincibles Studio

Invincibles Studio

- Preston based mobile games studio
- 100M+ users
- Focus on Football themed games
- 3 games in active development
- Key questions:
  - How should we expand?
  - How should we invest?

Timeline

2025

2024

2023

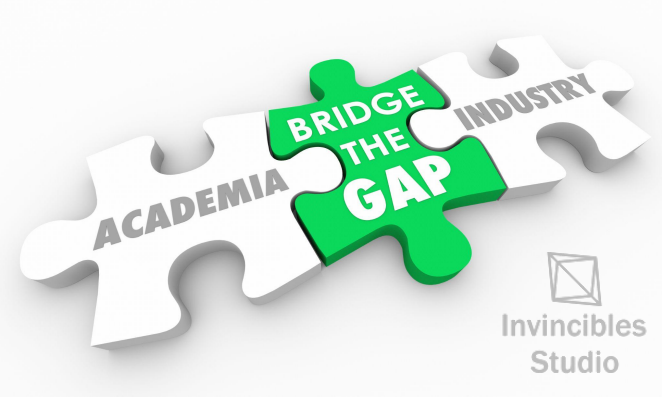
2022

2021

Invincibles Studio

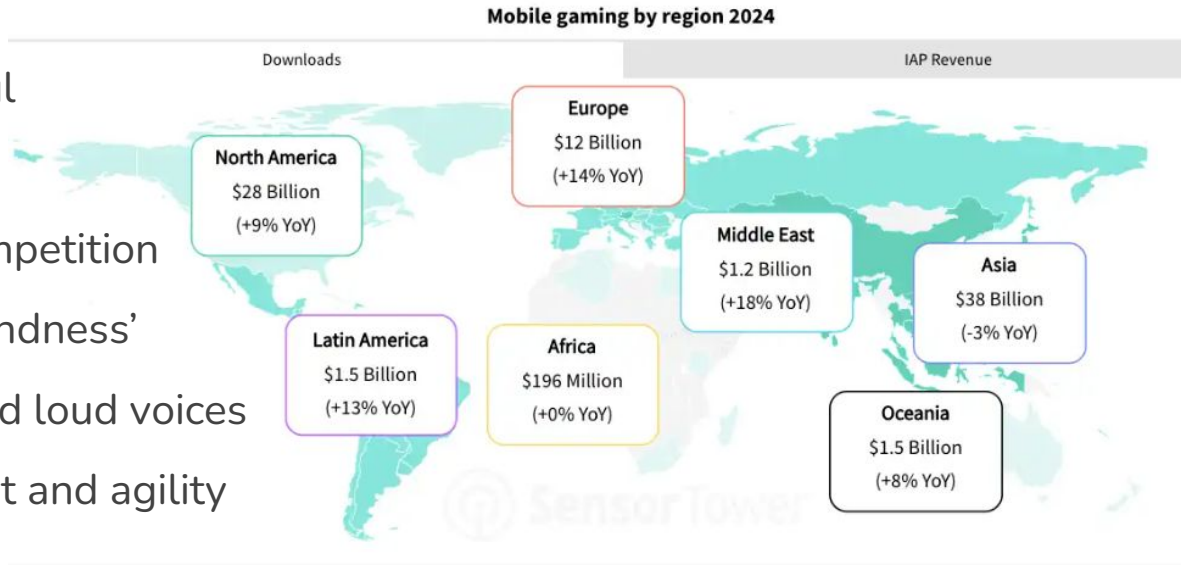
How Do Academics and Industry Link?

- Strong focus on achieving results
- Problem-solving skills matter more than specific knowledge
- Ability to plan and execute complex tasks without predefined solutions
- Visualisation skills
- Presentation skills
- Unique combination of all these abilities



Why Does Data Matter in Gaming?

- Data is cheap, and powerful
- Booming marketplace
- This means ever fiercer competition
- Overcoming ‘Developer Blindness’
- Customer segmentation and loud voices
- Enables faster development and agility
- Psychology meets data

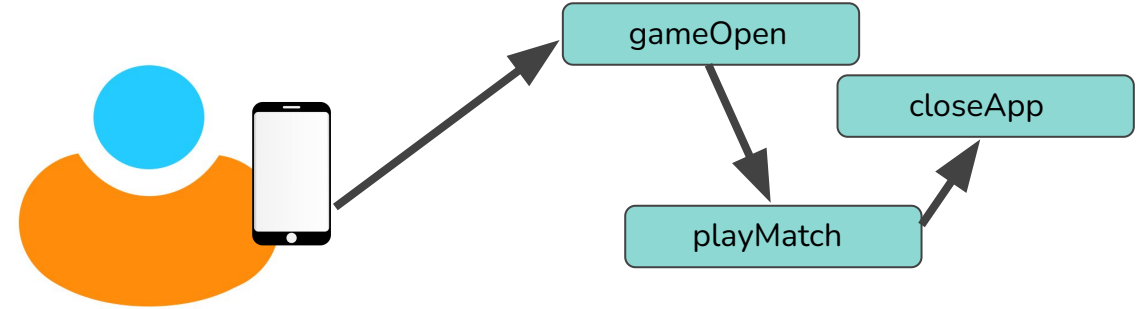


Credit: Sensor Tower

Invincibles Studio

How Does Data Work in Gaming?

- Typically the app industry uses “events based data”
- We create specific actions a user can perform, and log an event each time that happens
- This creates huge data sets (typically ~TBs per day)



UCLAN_example Run Save Download

```
1 SELECT event_name, event_timestamp
2 FROM `soccer-manager-2025.scheduledTables.baselineEventsTable`
3 WHERE table_date between DATE(2025,09,11) and DATE(2025,12,01)
4   and event_name = "user_engagement"
5   and user_id = "testUser"
6 ORDER BY event_timestamp asc
```

Logged and query-able in our data warehouse or through API.

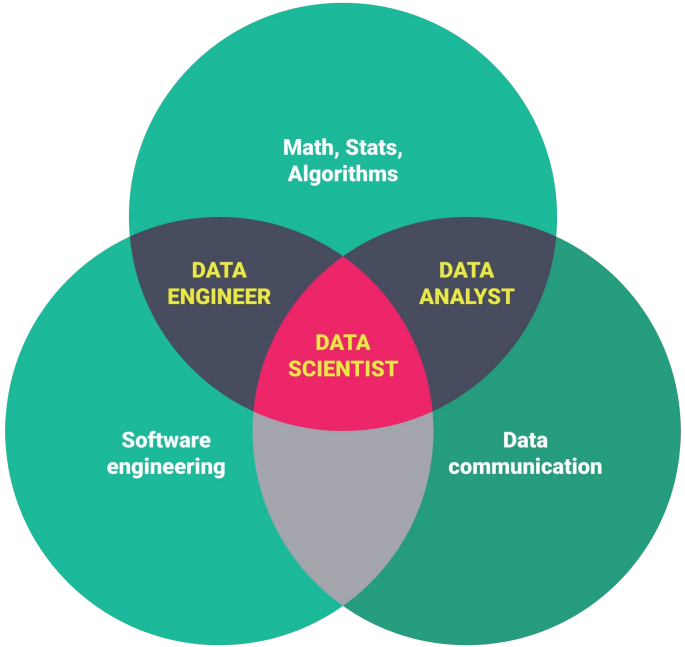
Invincibles Studio

The Roles of Data in Business

There are 3 key data roles:

- Data **Analyst** - someone who asks and answers questions with statistics
- Data **Scientist** - someone who creates models and predicts behaviours
- Data **Engineer** - the person who ensures the data flows

All 3 have large areas of overlap



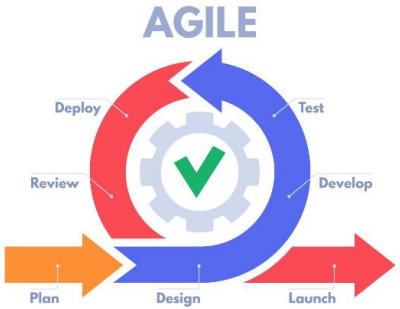
Credit: Writuparna Banerjee

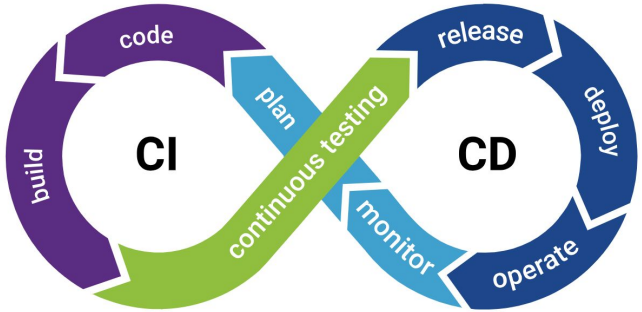
Invincibles Studio

Real World Examples

Typical Data Led Workflow

1. Ask a key question
2. Back of the envelope calculation
3. Acknowledge data sources
4. Model selection
5. Data munging
6. Acknowledge key statistics
7. Run test/create model
8. Investigate results
9. Report to key stakeholders





Invincibles
Studio

Data Analytics

Analytics

Analytics focusses on asking questions and producing statistics led results to answer them

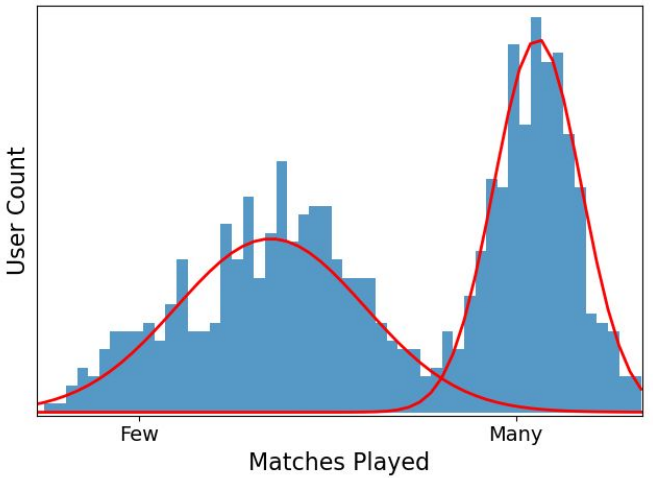
Activities include:

- Data fitting and uncertainty analysis
- Forecasting future values (e.g. SARIMA)
- A/B testing new features, assessing their impact
- Product stability analysis (e.g. tools like Crashlytics)
- Exploratory analysis: questions can be open ended - leads to unexpected insights

Invincibles Studio

Analytics - Example

- **Question**: SM21 - How do users play the game?
- **Process:**
  - Analyzed player behavior, accounting for biases like playtime
  - Identified two distinct player categories
  - Discovered a “road to glory” gameplay style
- **Outcome:**
  - There was potential for developing a completely new game mode to improve UX
  - Huge change, required careful AB testing



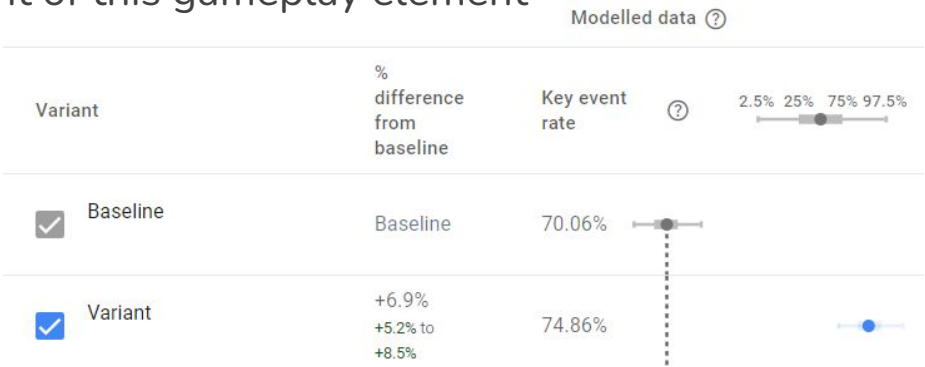
Resampled data set to clearly show differing cohorts playstyles, a clear bimodal distribution of matches played

Invincibles Studio

Analytics - Example

After developing the new game mode, we conducted further A/B testing to validate its value

- **Expected Impact:** modelled an anticipated 7% uplift in engagement and revenue
- **Testing Plan:** for necessary power, required a experimental period of 2 months
- **Results:** clear acceptance of the alternative hypothesis, we continued development of this gameplay element



A/B Test output, shows acceptance of alternative hypothesis (P<0.05)

Invincibles Studio

Analytics

Core Interests:

- Interested in data mining and exploring novel questions
- Creating impactful visualisations to communicate insights effectively
- Clear communication skills, tailored to different audiences

Technical Expertise:

- Tools: SQL, Tableau (Looker, PowerBI, and Excel-style tools)
- Programming: Python (Plotly, NumPy, polars) and R (ggplot2, dplyr)

Relevance to Gaming:

- Strong connection to the design and development elements of video games, integrating data-driven insights into gameplay and UX

Invincibles Studio

Data Science

Machine/Deep Learning

Data science is more development focussed, producing models to deliver customised experiences

Activities include:

- Being tasked by business intelligence or product managers to create custom predictive models
- Focus on key ML and DL techniques (Regression, Random Forest, Boosting, NN architecture)
- Creating custom user experiences

Data science involves *less* focus on investigation, more focus on delivering value

Invincibles Studio

Machine Learning - Example

Question: SM24 - How can we deliver impactful push notifications?

- **Challenge:** push notifications have a huge effect on retaining users, but people get bored/annoyed by them
- **Objective:** limit push notifications to only those that are absolutely necessary
- **Solution:** by creating a daily prediction of user **churn,** we can send a notification only when we believe a user is already lost
- **Approach:** binary classification problem, models such as logistic regression, random forest, boosting or NN are appropriate

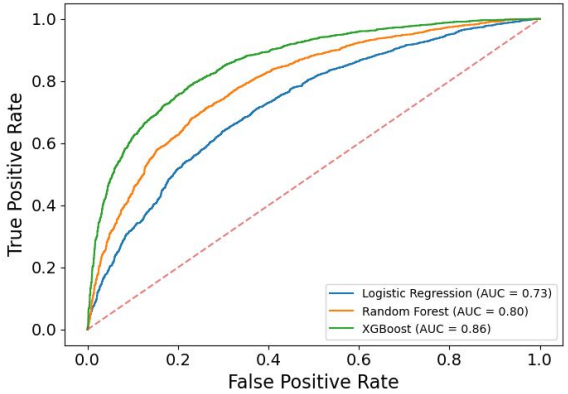
Invincibles Studio

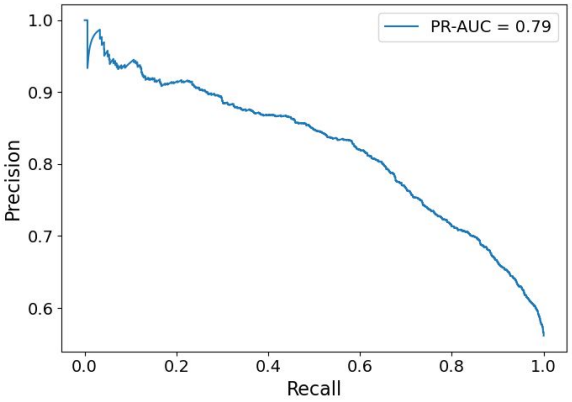
```
"""
Pre-processing tool for events based ML

@author: henry.childs_invinci
"""

import pandas as pd
import numpy as np
from sklearn.model_selection import train_t
import xgboost as xgb
from sklearn.tree import export_graphviz
from numba import jit
import random

default = {
    "starting_step": "startDataPack",
    "n_steps": 20,
    "events_per_step": 5
}

def filter_starting_step(x, starting_step="
    """
    Filter users who only begin at starting
    """

    starting_step_index = x.index(starting_
        if starting_step in x else 0

    return x[starting_step_index: starting_


@jit(nopython=True)
def user_journey(events, starting_step="sta
    """
    Map out user journey, filtering down du
    """

    events = events.sort_values(['user_id',

    # find users who complete starting step
    valid_ids = events[events['event_name']

    # create user journey
    flow = (
        events[events['user_id'].isin(valid
        .groupby('user_id')
        .event_name.agg(list)
        .to_frame()['event_name']
        .apply(lambda x: filter_starting_st
        .to_frame()['event_name']
        .apply(pd.Series)
    )

    flow = flow.fillna('End')

    # add the step number as prefix to each
    for i, col in enumerate(flow.columns):
        flow[col] = '{}: {}'.format(i + 1,
```

Machine Learning - Workflow

Model Selection and Development

- Define objective and success metrics
- Explore and clean data - training, validation and test sets
- Feature engineering
- Select candidate models and build
- Evaluate performance; ROC, PR, confusion matrix etc
- Iteration; hyperparameters, regularisation if needed
- Deploy and set up post monitoring

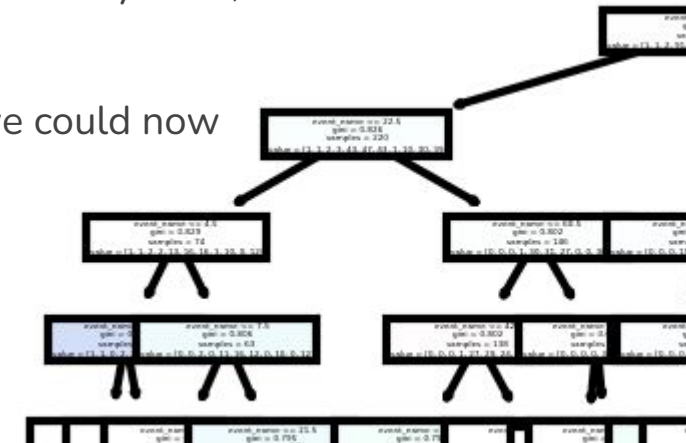




Model selection curves for churn prediction

Machine Learning - Example

- **Model Selection:** we chose XGBoost due to its strong predictive performance and ease of deployment within our cloud environment
- **Advantages:** XGBoost is robust, performing well on imbalanced datasets and feature-rich data
- **Performance:** We achieved a predictive accuracy of approximately 82%, which met our stakeholder requirements
- **Integration:** by integrating with other cloud based tools, we could now better serve push notifications to a user



Invincibles Studio

Leveraging Deep Learning

Neural networks are incredibly powerful: they can discover novel patterns and capture complex, highly non-linear relationships beyond traditional models.

They are, however, data hungry and time-consuming to iterate and deploy.

Best utilised with:

- Very large datasets
- Complex patterns
- High cardinality features
- Mix of structured and unstructured data

Invincibles
Studio

Data Science

Core Expertise:

- Understanding and applying model and feature selection techniques
- Collaborating with developers to integrate models into production systems

Technical Proficiency:

- **Python:** PyTorch, TensorFlow, Transformers, SciKit Learn
- **Cloud Services:** AWS, GCP, Azure (for deploying and managing models)
- **Prebuilt Models:** leveraging tools like Hugging Face, AI APIs (e.g. ChatGPT)

Focus:

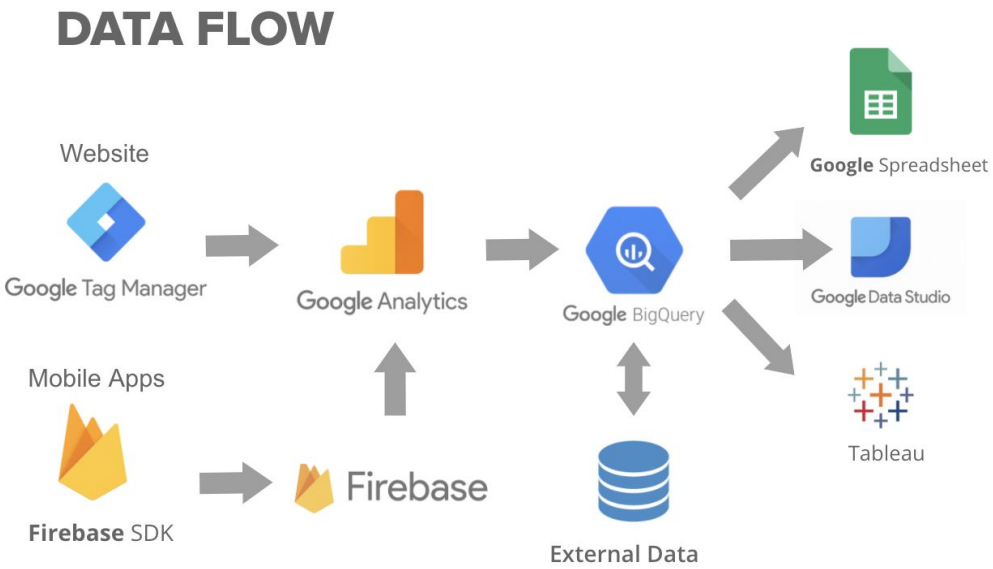
- Less emphasis on exploratory investigation
- More attention on delivering higher value through actionable, production-ready models

Invincibles Studio

Data Engineering

Data Engineering - Tech Stacks

Analytics and science rely on having quality data, but how do we get that?



Google Cloud

aws

Credit: Google

Invincibles Studio

Data Engineering - Data Quality

Data Quality Pillars:

- Completeness, accuracy, consistency and freshness

Data Dropouts:

- Missing events due to crashes, disconnects, SDK failures

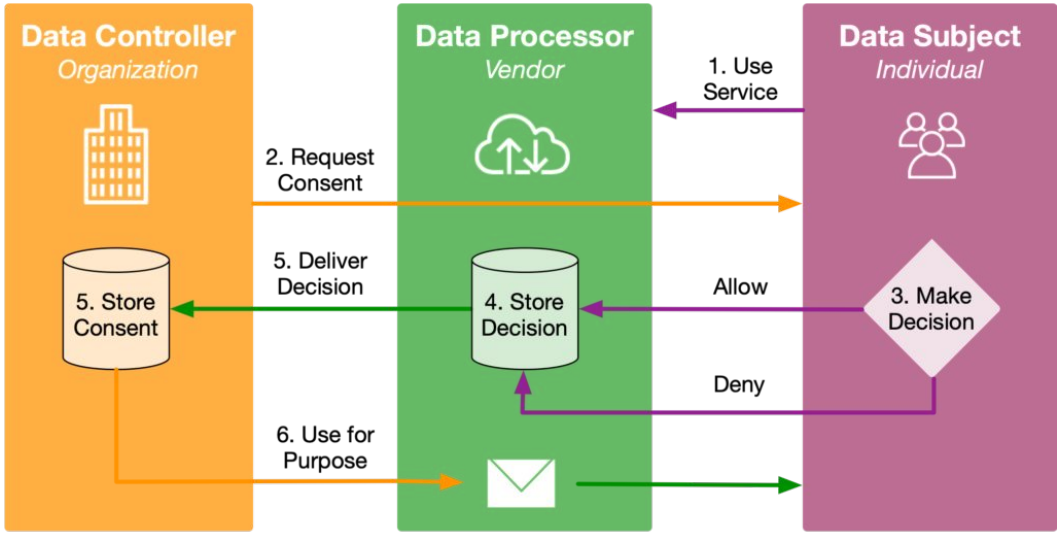
Data Cleaning:

- Remove corrupt/duplicate events
- Standardise formats
- Ensure timestamps are valid
- Impute/backfill where appropriate

Source: https://profisee.com/platform/data-quality/

Data Engineering - Privacy and Compliance

Evolution of big data has also increased the importance of safety and regulation



Example GDPR consent flow

Credit: Salesforce/Tableau

Invincibles Studio

Data Engineering

Core Technical Skills:

- Programming; Python, Scala, PHP
- Data pipeline creation (SQL, CICD, in-house and cloud)
- Cloud Services such as AWS, GCP and Azure

Core Competencies:

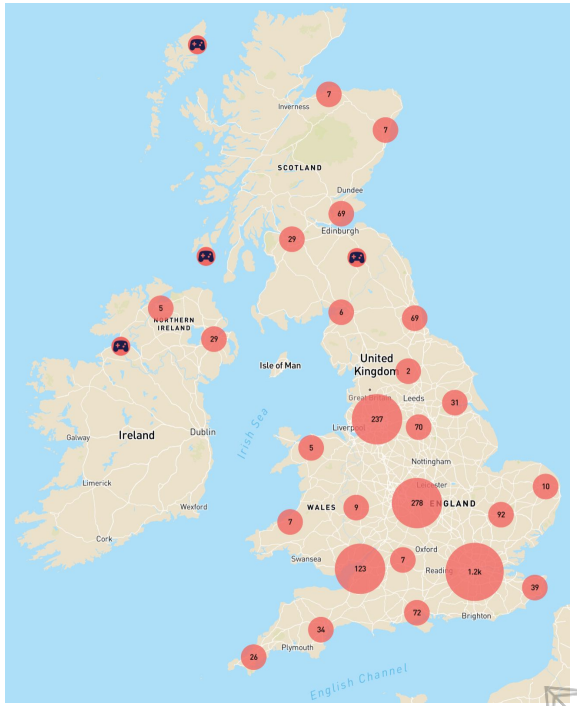
- Adaptation to new APIs and architectures, driven by development needs
- Strong interest in understanding and optimizing how data flows within an organization.
- Focus on **data velocity**
- Adhering to **data quality** principles - garbage in leads to garbage out

Invincibles Studio

Video Game Sector Opportunities

Now is an exciting time to pursue a data career within the gaming industry:

- Advanced tools and techniques are constantly emerging, the value of good quality data is consistently increasing
- Data is driving innovation, allowing for more creativity - e.g. data for improving UX
- Most game developers now have specialised data teams
- A rewarding career that allows you to directly see your work have measurable effect
- Use cutting edge tools and play with huge data sets



Credit: map.gamemap.uk

Invincibles Studio

Any Questions?

henry.childs@invinciblesstudio.com