**Video Game Localization Prioritization Tool**

Capstone Three Project Proposal - Joshua Ogden-Davis

**Problem statement**

We will build a tool to approximate a specific language market’s interest in a game based on its genre and price point. This tool could be used by game studios, publishers, or language service providers to instantly get a rough idea of the ROI for localizing a specific title into a specific language.

*The test case will be a Suspense/Horror game priced at the equivalent of $19.99, and we will attempt to reliably determine whether it will receive above average or below average interest from the Simplified Chinese market on Steam.*

**Context**

Localization is an expensive gamble. Even when using generative AI, a large amount of personnel hours goes into re-creating a game (and all its assets) into a new language. If the game does not fit the target audience’s interests well enough, there is a real danger that the publisher will not recoup the localization costs.

While larger publishers may have a general idea of which markets to localize into, their approach might not be data-driven. At smaller publishers or at self-publishing studios, international market analysis expertise may not exist at all. A standalone tool that could give an indication of a specific market’s interest in similar games could help these companies quickly estimate the relative success potential of their game in various languages.

**Criteria for Success**

The project will be successful if it can estimate a language market’s “Interest” in a game given the genre and price point as represented on Steam. Additional significant features may be discovered along the way.

“Interest” will be measured as a percentage of a game’s reviews that are in the target language minus the average number of reviews in that language for all games. This will give us an idea of whether a certain genre performs better or worse in that market than other genres.

**Scope of Solution Space**

The result will consist of a numerical indicator of the market’s relevant interest based on publicly available sales & review data.

**Constraints**

The relationship between “interest” and actual sales may not always be straightforward. We also will not know whether the relative percentage of reviews is a consistent indicator until the scraper is already built and implemented.

**Stakeholders**

The game publisher would provide the genres & price points to be tested against the model. Steam, as the holder of the key data, will play a role in determining the availability of the data.

**Key Data Sources**

The majority of our data will come from scraping the Steam platform. I have custom-built a scraper that has already collected 10k records.

If this data proves insufficient for making reliable predictions, we can explore integrating other data sources (such as the “Video Game Sales” [dataset](https://www.kaggle.com/datasets/gregorut/videogamesales) on Kaggle) to see if we can find relationships between a larger number of variables.