**Twitch Channel Sponsorship Tool**

**C964**

**Capstone**

**Jared Hines**

# Table of Contents

Letter of Transmittal 4

Project Proposal 5

Problem Summary 5

Data Product Benefit 5

Data Product Outline 5

Data Description 6

Objective and Hypothesis 6

Methodology 6

Funding Requirements 7

Solution Impact 7

Sensitive Data Precautions 7

Expertise 7

Executive Summary 8

Problem Statement 8

Customer Summary 8

Existing System 8

Data 9

Methodology 9

Deliverables 10

Implementation 10

Validation 10

Resource Costs 10

Development Schedule 11

Documentation 12

Business Vision 12

Dataset 12

Descriptive Methods 14

Predictive Method 16

Hypothesis Verification 17

Effective Visualization 17

Accuracy 17

Product Testing 17

# Table of Contents Cont.

Source Code 18

Quick Install Guide 19

Sources 20

# **Letter of Transmittal**

Jared Hines

Systems Developer

Sponsorship Connecting Company

4001 S 700 E #300

Millcreek, UT 84107

Dear Mr. Bossman,

We have noticed that your company has shown interest in sponsoring Twitch channels to drive growth and profits. Twitch is a large company with plenty of opportunity to showcase some of your products on some channels. An issue that may arise when sponsoring a channel is sponsoring the right candidate who could be a good brand ambassador for your company and products. We have a solution for you, a tool that gives valuable data insight to top channels to give you more information about who you would want to sponsor. Sponsoring the right channels will grow your company.

The proposed tool would include an interactive environment to pull data and visualize it the way that you would want to see it. It would allow you to look deeper into the data and see trends and distributions that are not commonly known. With little training, your team can manipulate the data to see it in a way that can drive more growth to your company. Many channels may be brand risks or may seem bigger or smaller than they actually are. Sponsoring the right channel to advertise your company and products is imperative to the success of your company.

The proposed tool would come at a cost of $20,000. A large portion of this money would be used for resources and team costs. We have a team of experienced individuals that will ensure that the product satisfies needs that you may have. They are also very reliable. A training on how to use the tool will be included in the purchase.

Please reach out to me if you have any questions. If you want to schedule a meeting to discuss the product, send me an email so we can discuss an agenda and location. I am looking forward to hearing back from you.

Sincerely,

Jared Hines

Systems Developer

# **Project Proposal**

## Problem Summary

Sponsorship Connecting Company is an entity that wants to connect businesses with channels to help grow business partnerships. Many companies do not know which channels they want to sponsor or how many people would view their advertisements. Channels have different followings, languages, mature ratings, brand risk, average viewership, and much more. The proposed tool will allow insight to metrics that aren’t easily found by a simple search on the internet. The tool will allow companies to look at data and allow them to determine what is important to them and then decide who they want to sponsor.

## Data Product Benefit

Sponsorship Connecting Company has reached out to several companies and asked about their process to choose channels to sponsor. The current trend is to see who has the most viewers when they visit Twitch. Although this does seem useful, it is unlikely that it is the best system to use. Our tool will be utilized to access trends and distributions using real data. The user can generate data visualization methods in order to have a graphical understanding of the data. The tool also includes a regression analysis that will allow companies to choose what metrics are important to them and see how channels compare to their values. The overarching benefit of the tool would be a great business partnership between sponsor and channel.

## Data Product Outline

The data product was built with Python. Libraries that are useful to the tool for data analysis and visualization were also added. The product was built within a Jupyter Notebook. The product will also be run as a Jupyter Notebook so that the consumer of the product can manipulate the variables, determine what graphics they would like to use, and perform regression analysis on their own. All of these have will be coded into the tool already. There will be a training guide and support from our company to show how the tool can be utilized for maximum potential. Included in the tool is a value search, channel search, and a way to find if data is missing.

## Data Description

Data used in the tool was retrieved from Kaggle.com. Kaggle is a website that has many types of datasets available for use. The dataset for the tool contains the information of the top 1000 Twitch channels from 2020. The data is in CSV format so that it can be easily read by the program. The proposed solution will scale to much larger databases and will be able to integrate new data easily. Data that is included has these values:

* Channel (string)
* Watch Time (integer) \*Note – value shown in minutes
* Stream Time (integer) \*Note - value shown in minutes
* Peak Viewers (integer)
* Average viewers (integer)
* Followers(integer)
* Followers Gained (integer)
* Views Gained (integer)
* Partnered (Boolean)
* Mature (Boolean)
* Language (string)

## Objective and Hypothesis

The objective of the tool is to allow companies to find channels that they believe will benefit them from sponsoring. Companies will use the data given to them to determine which channels might be the best fit to sponsor. The tool will be used to connect businesses with Twitch channels indirectly. If the tool is utilized correctly by a company, then they will see a large growth in business partnerships and profit through their advertising and sponsorships.

## Methodology

This project will use agile methodology. Due to the nature of Jupyter notebook, different areas of the notebook can be tested individually. The code does not have to be run sequentially. Agile methodology requires constant communication and preparation. Testing can be taken place during some phases of development if needed. Use of agile methodology will allow for quicker production of the tool. The process is broken down into five phases. The phases are requirements, development, testing, delivery, and feedback. Requirements phase deals with getting the information from the customer on what they need the product to do. Development will code the product. Testing will make sure that the product runs as intended. Delivery will give the finished product to the customer. Feedback will allow the customer to let the developer know what the product can do better.

## Funding Requirements

The initial cost of the project will be $20,000. This includes cost of resources and pay for the development team. If there are additional features to be added at the request of the customer, more funding will be required. Funding of additional requirements will be assessed at the time of the request.

## Solution Impact

The impact of the solution on stakeholders will be significant. The tool will allow the company to grow tenfold. Stakeholders will see value in the form of business partnerships and profits.

## Sensitive Data Precautions

This tool will not be using any sensitive data. The data is all publicly available. In the event that a customer wants to keep their data confidential, it will be feasible due to the tool being local. The tool will be locally run on a server while also using locally stored data. All information shared by the companies after the tool is given to them is to be at their discretion. It is recommended that all files are encrypted, and all communications are secure.

## Expertise

The development team is a group of 4 experienced programmers who specialize in tools that help visualize data. They have worked for many years using the agile methodology and will get the work done quickly and professionally. This team of developers will create a great, usable product in a timely manner.

# Executive Summary

## Problem Statement

The Twitch Channel Sponsorship tool that we will develop is going to increase the amount of data available to a user, while reducing the amount of time and research needed to find a good candidate. Data visualization and easy to access/manipulate code will be available for more processes to be ran. Graphs will be easily drawn from data and used for business purposes. Productivity will be increased, and more value will be seen from the business and its stakeholders when this tool is implemented.

## Customer Summary

The Twitch Channel Sponsorship tool will be used by companies who are looking to sponsor channels. Through sponsorship, the companies will be able to advertise their products or services to a large audience. The usefulness of their products and services will also be seen because the channels will utilize their products and services, and their viewers would be more inclined to purchase the products. This tool will increase the size of the company, while also reducing load for marketing and sponsorship.

## Existing System

The current system used by most companies is by searching for top streamers without looking at the data. This search will typically output who is currently live with the most viewers or followers. This may not give the full picture of the actual data behind the channels. A lot of the sponsorships are reached out to by talent organizations that do not have full datasets, instead they have clients they are obligated to reach inform about the sponsorship even if they are not a good fit. This system may also lead to large channels getting sponsorships that are not the right demographic, language, or the channel may not be reaching enough viewers. As of now, there is not a system in place, so our product will solve that issue.

## Data

The current dataset is the top 1000 channels from Twitch. The dataset was collected and published on Kaggle.com. The data is available as listed:

* Channel (string)
* Watch Time (integer) \*Note – value shown in minutes
* Stream Time (integer) \*Note - value shown in minutes
* Peak Viewers (integer)
* Average viewers (integer)
* Followers(integer)
* Followers Gained (integer)
* Views Gained (integer)
* Partnered (Boolean)
* Mature (Boolean)
* Language (string)

The entire dataset will be utilized to ensure that testing goes smoothly. This data can also be used to begin making business decisions. Eventually, new datasets will come out and can replace this to be entered into a csv file for upload to the sponsorship tool. Data should always be in .csv files so we can ensure that they are compatible with the product. A link to this dataset is placed below.

<https://www.kaggle.com/datasets/aayushmishra1512/twitchdata>

## Methodology

Agile methodology will be the selected production method. This allows for certain parts of the project to proceed without waiting for the project as a whole. There will be constant communication between the customer, developers, and project managers. The first part of the methodology is the requirements phase. During this phase, requirements made by the consumer will be communicated to the entire development team. Requirements set the minimum standards for what needs to be added to the project. The next phase is the development phase. During this phase, developers will create the system. The third phase is testing. During this phase, testing will take place by a software quality assurance team to ensure that product is working properly. The fourth phase is delivery. The product will be delivered to the customer. If the customer is not satisfied with the product, then requirements will be reviewed, and we will go through the process again. Once the customer accepts the product, the feedback phase will begin. Feedback from the customer will take place. If more features are needed, requirements change, and funding will need to change. Once new requirements are set, the method runs again. These steps will repeat until the customer is satisfied.

## Deliverables

The deliverables associated with this project will be the Jupyter notebook tool and the data associated with the tool. Things also to be shared with the customer include the schedule, budget, documentation, and training/instructions for the tool.

## Implementation

Using agile methodology, we plan to develop and deliver a product in a timely manner. In order to do that, constant communication must ensue from the beginning. Clear goals and objectives will need to be stated and met. Once the product is believed to be complete, quality assurance will test the product. Assuming the product pasts those tests, we will have random testing by third party users who do not have stake in the product. Bugs and errors will be fixed and tested until the project runs smoothly. Once the testing has completed and requirements are met, it will be sent to the customer for acceptance testing. At the end of everything, a fully functional product will be delivered, and feedback will be provided by the company.

## Validation

To ensure that the product meets the requirements of the customer, in house and third-party testing will take place. We will use quality assurance testers to ensure that basic and advanced functions are error free. Once testing is deemed satisfactory by testers, the product will be sent to the customer to determine if it will be accepted or not. If it is not accepted due to requirements changing, more funds will be required.

## Resource Costs

The product will be developed using Python 3.11.0. Libraries within Python will also be utilized. Jupyter notebook is the programming environment. All are free to use and can be used on most operating systems. The product will be able to run locally, and server cost will not be present. It will be developed using Windows 10. Total Cost = $0

The initial cost of the product is $20,000. All of the following includes human hours cost, resource cost, and other expenses. The cost breakdown includes initial development and testing the system. This is phase 1 and will cost about $7,500. The next part of the project is creating the fully functional project that can be implemented in a business setting. This is phase 2 and will cost $7,500. The final part of the project will be acceptance testing and feedback. This is when the product will be tested by the company and final feedback is given. Support by the development team is given to the customer throughout acceptance testing. This is phase 3 and will cost $5,000. That ends up being a total of $20,000.

## Development Schedule

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Event** | **Milestone** | **Start** | **End** | **Duration** | **Dependencies** | **Resources** | **Phase** |
| 1 | Requirements meeting | 11/1/22 | 11/1/22 | 1 Day | None | Development Team, Stakeholders, QA Team | - |
| 2 | Project start meeting | 11/3/22 | 11/3/22 | 1 Day | 1 | Development Team, Stakeholder, QA Team | - |
| 3 | Project planning and design | 11/4/22 | 11/5/22 | 2 Days | 2 | Development Team | 1 |
| 4 | Data review | 11/6/22 | 11/6/22 | 1 Day | 3 | Development Team | 1 |
| 5 | Project Development | 11/7/22 | 11/12/22 | 5 Days | 4 | Development Team | 1 |
| 6 | Project testing | 11/12/22 | 11/14/22 | 2 Days | 5 | QA Team | 1 |
| 7 | Deliver initial product | 11/15/22 | 11/17/22 | 2 Days | 6 | Development team, Stakeholders | 1 |
| 8 | Feedback meeting and planning | 11/18/22 | 11/18/22 | 1 Day | 7 | Development team, stakeholders, QA team | 1 |
| 9 | Project development | 11/19/22 | 11/23/22 | 4 Days | 8 | Development team | 2 |
| 10 | Testing | 11/24/22 | 11/25/22 | 2 Days | 9 | QA Team, 3rd party testers | 2 |
| 11 | Deliver product | 11/26/22 | 11/26/22 | 1 Day | 10 | Development team, stakeholders | 2 |
| 12 | Acceptance testing | 11/27/22 | 11/27/22 | 1 Day | 11 | Stakeholders, Dev Team, QA team | 3 |
| 13 | Product Delivery | 11/28/22 | 11/28/22 | 1 Day | 12 | Stakeholders, Development team, QA team | 3 |

# Documentation

## Business Vision

The business vision is to have a product that delivers data and data visualization to the user. The user interface is a Jupyter notebook because it is interactive. It can report missing data in a column, draw visualizations, calculate regression, and other functions. The end user will see data delivered to them for use in business practices.

## Dataset

The data set has already been cleaned by someone else before it was retrieved from the internet. A picture below shows the data in the csv file.

Table

Description automatically generated

Picture of Dataset

Below is a picture of the code used to detect any empty data in the set.

Graphical user interface, text, application, email

Description automatically generated

## Descriptive Methods

Below are pictures of the code used for the descriptive methods. An example of each one is also showed.

### PieGraphical user interface, text, application, email Description automatically generated

Chart

Description automatically generated

### Bar

Graphical user interface, text, application, email

Description automatically generatedChart

Description automatically generated

## Predictive Method

A linear regression model was constructed to try to predict where a channel’s value should be based on two variables. In the tool, you can hover your cursor above the line to see the predicted values based on the data. Descriptions of how the models can be interpreted are also within the tool. The code and an example are shown below. Graphical user interface, text, application

Description automatically generatedChart, scatter chart

Description automatically generated

## Hypothesis Verification

The hypothesis can not be verified at the current moment. The initial hypothesis was that a company would see growth if the tool was used correctly. As of now, the tool provides the user with the insight and information they need to see if they should sponsor a channel. The tool works as intended so now the company must use that information to grow their business. It seems that the progress with the tool and work with the company has moved it in the right direction, not confirming but supporting the hypothesis.

## Effective Visualization

As shown before, the visualizations are present in the tool. Included in the tool are pie charts, bar charts, and regression graphs. The Jupyter notebook is interactive so with a little bit of training and instruction, a simple line of code can be added to create a new visualization. This gives the user the freedom to visualize the information they want to use, not just the information we provide.

## Accuracy

The accuracy of the tool can be first shown by the completed data set function. By knowing that all values are present, we can ensure that the linear regression method will run smoothly. Also, a channel search is included to look up values of certain channels. The search returns the entire row of values and can be checked against the dataset to ensure that it is correct. Because all values are present and all data is entered in the corresponding columns, the visualizations are all accurate representations of the data.

## Product Testing

During testing, we ran into an issue where some of the libraries where not loading or installing correctly. A function was put into place to except the error and make sure that the library was imported and installed. Markdown text was also added to the tool for easy explanation of code and how to manually change it through the interface. Markdown text was also added for organization. Text, letter

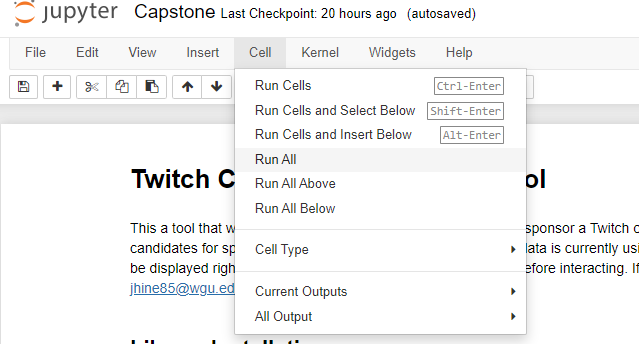
Description automatically generated

## Source Code

The source code is located in a file called Capstone.ipynb. The dataset is in a file called twitchdata-update.csv. These two files are all you will need for the tool.

## Quick Install Guide

1. Install Python from the download menu on python.org
2. Install Jupyter Notebooks using the installation guide from <https://jupyter.org/install>.
3. Download C964.zip
4. Extract the file and place it on the desktop
5. Open the folder location
6. Type “cmd” into the address bar as shown in the picture belowGraphical user interface, text, application, email

   Description automatically generated
7. Type “jupyter notebook”
8. Press ENTER
9. If you have used Jupyter notebooks before, it should launch automatically. If you have not, select your preferred browser to open the file.
10. Click on Capstone.ipynb
11. Click on Cell on the top bar and select Run All
12. You can now freely explore the page and data visualizations. You can also add code if you please.

# Sources

Data obtained from Kaggle.com

Mishra, Aayush. “Top Streamers on Twitch.” *Kaggle*, 24 Aug. 2020, https://www.kaggle.com/datasets/aayushmishra1512/twitchdata.