# Project Documentation: CrowdWisdomTrading AI Agent Approach

## 1. Approach Overview

The solution to the CrowdWisdomTrading assessment is a backend Python script built on the CrewAI framework. The core objective is to create a multi-agent system that unifies product data from various gambling and prediction market websites. The workflow is designed to be a sequential pipeline where three specialized agents collaborate to perform data collection, analysis, and final report generation. The entire process is automated, ensuring a seamless flow from raw data scraping to the final structured CSV output.

## 2. Agent Design

The system is composed of three distinct agents, each with a specific role, goal, and backstory to ensure a focused and effective workflow.

### Agent 1: Data Collector

Role: Web Scraper and Data Acquisition Specialist

Goal: To efficiently and reliably scrape raw product data from specified gambling and prediction market websites, providing the output in a clean JSON format.

Backstory: You are a highly skilled and persistent data acquisition agent. Your expertise lies in navigating complex web structures to extract specific information. You are known for your precision and thoroughness, ensuring that no relevant data is missed. Your primary objective is to lay the foundation for the entire analysis by providing a comprehensive and accurately scraped dataset.

### Agent 2: Data Analyst

Role: Product Data Analyst and Identifier

Goal: To process collected JSON data, analyze product descriptions and market details, and accurately identify which products are identical across different websites, providing a structured JSON output.

Backstory: You are a meticulous analyst with a keen eye for detail. Your specialty is in pattern recognition and data normalization. You can discern subtle differences and similarities between products, even when their names or descriptions vary slightly across different sources. Your job is crucial for creating a unified view of the data, ensuring that the final report is based on correctly matched products.

### Agent 3: Data Formatter

Role: Data Formatter and Reporting Analyst

Goal: To transform the analyzed JSON data into a well-structured and organized CSV file, including unified product lists, their prices from various sites, and a confidence level for matching.

Backstory: You are an expert in data presentation and final reporting. Your mission is to take the processed, raw data and format it into a clear, concise, and easy-to-read CSV file. You understand that the final output must be user-friendly and ready for immediate use, making complex information accessible. You are the last step in the workflow, ensuring all previous efforts are presented in a professional and actionable format.

## 3. Task Flow and Execution

The workflow is a sequential chain of tasks, where the output of one task serves as the input for the next. This ensures a logical progression from data acquisition to the final deliverable.

### Task 1: Data Collection (data\_collection\_task)

Description: The first task is to navigate to at least three gambling and prediction market websites (e.g., polymarket.com, kalshi.com) and scrape all available product listings, including market name, current price, and any other relevant details.

Expected Output: A JSON object containing the raw, unprocessed data scraped from all target websites. The data from each site should be clearly separated within the JSON structure.

Agent: X Data Collector

### Task 2: Product analysis (product\_analysis\_task)

Description: This task analyzes the raw JSON data provided by the X Data Collector. It identifies and cross-references products that are the same across the different websites. The agent will normalize product names and create a unified list, noting which products correspond to each entry.

Expected Output: A JSON object representing a unified product list. Each entry in the list contains a standardized product name and a sub-object or array listing the corresponding product details (prices, original names) from each of the source websites.

Agent: Identify Products and If They Are the Same or Not

### Task 3: data formatting (data\_formatting\_task)

Description: The final task is to take the unified JSON product data from the previous agent and format it into a CSV file. The CSV must have the following columns: "Unified Product Name," "Price on Site 1," "Price on Site 2," "Price on Site 3," and "Confidence Level." The confidence level is a metric calculated based on the similarity of the matched products.

Expected Output: A CSV file named "unified\_products.csv" containing the final, organized product data as specified. The file should be ready for viewing and sharing.

Agent: Re-arrange the Unified Data in a Great UI

## 4. Overall Workflow

The CrewAI workflow will be initiated with a kickoff call that triggers the data\_collection\_task. This task will execute and pass its JSON output to the next task in the sequence, data\_analysis\_task. The second task will perform its analysis and, upon completion, pass its new JSON output to the final task, data\_formatting \_task. The third and final task will then use this data to create and save the final.csv file, completing the entire process. This chained execution ensures a clean, modular, and efficient pipeline for data processing.

Note: As I was using an open source model, I wasn’t able to properly complete this flow as the tokens limits were exceeding but the workflow is designed to work well with better paid models so try running it with openAi GPT-4 api key (paid) instead of Groq llms api if the workflow stops in between or facing any internal server errors.