**About Tezda**

**Company Overview**

Tezda is an **innovative startup** based in **London**. Our team is dedicated to building **immersive social experiences** that harness cutting-edge technologies—such as **machine learning (ML)**, **short-form video platforms**, **blockchain**, and **Web 3.0** principles—to **redefine how people connect, create, and transact online**.

Originally grounded in e-commerce and blockchain solutions, we’ve evolved into a **Social Hub platform** that merges the interactivity of short-form video (think TikTok or Instagram Reels) with **immersive, community-oriented experiences**. Whether users want to discover new content, join thematic virtual hubs, or interact with creators and brands in real time, Tezda’s platform aims to be the go-to space for engaging, data-driven social interactions.

**Mission & Vision**

**Mission**

Our mission is to **empower users** through immersive social hubs, short-video content, and advanced **AI-driven recommendations**—creating a **personalized, interactive** environment that goes far beyond static scrolling. We believe in:

1. **Community & Connection**
   * Building hubs where users can engage in live sessions, short videos, or micro-communities that reflect their interests and passions.
2. **Immersive Engagement**
   * Leveraging **VR, AR**, and **3D** technologies to bring digital spaces and video content to life.
3. **Decentralized Ownership & Control**
   * Enabling secure, user-centric data control and **trustless transactions** via **blockchain** protocols and **Web 3.0** frameworks.

**Vision**

We envision a future where **digital social experiences** are as vivid and interactive as physical gatherings—where **short videos, virtual spaces, and community hubs** merge seamlessly to form a dynamic **metaverse**:

* **Seamless Content Discovery**: Users explore endless short video feeds tailored to their tastes by advanced **ML recommendations**.
* **Real-Time Interaction**: Instant communication, live streaming, co-watching, and VR meetups unify people across the globe.
* **Full Control & Trust**: Users own their data and identity, forging more **authentic** and **equitable** interactions in a decentralized ecosystem.

**Why Social Hubs?**

1. **Interactive Communities**:  
   Just like popular short-video apps, Tezda fosters connections through relatable content, creative expression, and real-time engagement.
2. **Immersive Multimedia**:  
   We go beyond static feeds, merging **AR/VR** experiences with short videos, so users can “step into” digital gatherings, product demos, or creative collaborations.
3. **Personalized Recommendations**:  
   Our platform uses advanced **machine learning** and **recommendation models** to continuously adapt, making content discovery both **fun** and **highly relevant**.
4. **Decentralized Transparency**:  
   By integrating **blockchain** features, we uphold transparency, trust, and ownership of digital assets—empowering communities to create and share value securely.

**Location & Team**

Our **London-based** office, is a vibrant hub of innovators passionate about pushing the boundaries of **immersive tech**. We value open collaboration, continuous learning, and a fearless approach to solving big challenges.

**Main Technologies**

1. **Machine Learning & Data-Driven Recommenders**
   * We build AI models (e.g., advanced recommendation engines, computer vision) to personalize user experiences and suggestion algorithms.
2. **Short-Form Video Platforms**
   * Short videos form a core interactive layer, allowing users to create, share, and discover bite-sized content in real time.
3. **VR & AR**
   * Virtual Reality (VR) and Augmented Reality (AR) bring experiences to life by immersing users in 3D environments or overlaying digital elements onto their physical surroundings.
4. **Blockchain / Web 3.0**
   * Our decentralized layer offers secure transactions, digital asset ownership, and user-governed economies that foster trust and collaboration.
5. **3D & Metaverse Elements**
   * We’re evolving toward an interconnected metaverse where short videos, live interactions, and digital commerce seamlessly converge.

**Opportunity for Candidates**

At Tezda, we’re seeking **talented individuals** who share our excitement for **next-generation social interaction**, **immersive content**, and **decentralized systems**. Whether your expertise is in **machine learning**, **AR/VR development**, **backend/blockchain engineering**, or **product design**, we’d love to hear from you if:

* **You Thrive in a Startup Environment**:  
  You’re adaptable, self-driven, and ready to solve challenges quickly and creatively.
* **You Value Collaboration & Openness**:  
  We believe in idea-sharing, constructive feedback, and pushing boundaries together.
* **You’re Passionate About Innovation**:  
  You keep up with the latest in ML, immersive tech, or Web 3.0, and love to experiment.
* **You’re User-Focused**:  
  You care deeply about crafting delightful user experiences and building solutions that empower communities.

By joining Tezda, you’ll help shape the **future of immersive social hubs**, gain hands-on experience with cutting-edge technologies, and **contribute to a platform** poised to transform how people **discover, engage with, and share digital content**.

**Join Us**

If our mission and vision resonate with you, we invite you to become part of our journey. Together, we can **redefine the social experience**—merging short-form video, immersive technology, and decentralized ownership into an exciting new paradigm. **Welcome to Tezda**, where innovation meets community in the heart of London.

**Technical Task: Movie Data Analysis & Recommendation**

**Dataset Description**

You have a CSV dataset (attached to this email) of ~4,800 movies with columns like:

* **index**
* **budget**
* **genres**
* **homepage**
* **id**
* **keywords**
* **original\_language**
* **original\_title**
* **overview**
* **popularity**
* **production\_companies**
* **production\_countries**
* **release\_date**
* **revenue**
* **runtime**
* **spoken\_languages**
* **status**
* **tagline**
* **title**
* **vote\_average**
* **vote\_count**
* **cast**
* **crew**
* **director**

An **example** row is shown above. The data includes both numeric (e.g., budget, revenue) and textual fields (genres, keywords, overview, etc.), as well as nested-like JSON structures for cast, crew, and production\_companies.

**Task Overview**

1. **Exploratory Data Analysis (EDA)**
2. **Data Cleaning & Feature Engineering**
3. **Recommendation Prototype**
4. **Dashboard or Visualization**
5. **Short Write-up / Documentation**

**Time Estimate**: ~4–5 hours

**1. Exploratory Data Analysis (EDA)**

1. **Data Loading & Inspection**
   * Load the CSV into pandas (or your preferred environment).
   * Show the shape of the dataset, column types, and any anomalies.
2. **Descriptive Statistics**
   * Basic stats: average budget, average revenue, distribution of vote\_average or runtime.
   * Identify the **top 10** highest-grossing movies, or highest budget.
   * Look for relationships between numeric columns (e.g., correlation between budget and revenue).
3. **Genre & Language Analysis**
   * Which genres are most common?
   * Which languages are most common? (e.g., count how often original\_language = “en”, “fr”, etc.)
4. **Visualizations**
   * Create **1–2 plots** (e.g., bar chart of top genres, scatter plot of budget vs. revenue).

**Deliverable**:

* A **notebook or script** showing your data loading, cleaning notes, and summary statistics/visualizations.

**2. Data Cleaning & Feature Engineering**

1. **Handle Missing or Irrelevant Data**
   * Check for NaN values or unexpected data in columns like genres, cast, budget, etc.
   * Decide if you drop rows/columns or fill them with placeholders.
2. **Parsing / Transforming Fields**
   * For textual fields like genres or keywords (which might look like JSON), parse them into a more workable format (e.g., lists of genre names).
   * You may also parse out the main cast or the director from crew.
3. **Feature Creation**
   * For recommendation or analysis, you might create a “combined features” column that merges title, keywords, cast, or genres.
   * Consider normalizing numeric features (e.g., budget, popularity) if needed for modeling.

**Deliverable**:

* Show how you’ve **cleaned** the dataset and **transformed** columns for better usage.

**3. Recommendation Prototype**

Using the cleaned dataset, build a **simple recommendation system**. Possible approaches:

1. **Content-Based Filtering**
   * Construct a similarity matrix (e.g., using TF-IDF on overview + genres + keywords) or parse cast/crew for similarity.
   * For a given movie “M”, find the top 5–10 “most similar” movies.
2. **Popularity or Hybrid**
   * Alternatively, recommend the **top N** movies by vote\_average or popularity in the same genre.
   * Or combine both “similar content” + “popular in the same genre.”

**Deliverable**:

* A **function** (e.g., recommend\_movies(movie\_title)) that outputs a list of recommended titles.
* Briefly explain the logic in your code or README.

**4. Dashboard or Visualization**

Create a **small dashboard or set of visualizations** (using any tool you prefer: **Tableau Public**, **Power BI**, **Streamlit**, or a simple **Matplotlib/Seaborn** interactive notebook). Show key insights such as:

* Top 10 movies by revenue or vote\_average.
* Distribution of genres or languages.
* (Optional) Interactive filter for year or runtime.

**If you choose Tableau/Power BI**:

* Export your final cleaned CSV or DataFrame, load into Tableau/Power BI.
* Publish your dashboard (Tableau Public link, or .pbix file).

**Deliverable**:

* A **link** (for Tableau Public) or **screenshots** of your dashboard.
* At least 2–3 relevant charts or interactive visuals.

**5. Short Write-up / Documentation**

In a **README.md** or equivalent:

* **Approach**: Summarize your data cleaning, EDA, and recommendation steps.
* **Challenges & Insights**: Note any interesting or surprising findings about the data (e.g., big outliers in budget).
* **Instructions**:
  + How to run your code (Python version, libraries, etc.).
  + If using a dashboard tool, include the link or file to open.
  + If you’ve built a function recommend\_movies(), show an example of how to call it.

**Submission Guidelines**

1. **GitHub / Zip Archive**
   * Include your code/notebooks, final cleaned dataset (or script for cleaning), and README.md.
2. **Dashboard Link** (Optional but recommended if using Tableau Public).
3. **Short Summary**:
   * In your submission email or comment, outline your main solution approach and any relevant instructions.

**Evaluation Criteria**

1. **Data Analysis & Cleaning**
   * How well did you handle missing/dirty data and parse fields like genres or cast?
2. **Recommendation Logic**
   * The clarity and correctness of your approach (even if basic).
   * Code readability and how well you explain it.
3. **Visualizations / Dashboard**
   * Quality and relevance of insights in your charts.
4. **Documentation & Structure**
   * Is it easy to follow your methodology?
   * Can we replicate your environment and results with minimal friction?
5. **Completeness**
   * Enough detail to demonstrate your skills without going beyond ~4–5 hours.
   * Clean, simple, and insightful solutions are preferred over complex but unfinished ones.