**Project Overview & Objectives**

The project aims to automate YouTube market analyses through a web-based platform equipped with an intuitive chat interface. The primary objectives are:

* Automate end-to-end workflows for niche market analysis (e.g., cricket channels, Turkish dubbed dramas).
* Provide actionable insights through visualizations and downloadable reports.
* Enhance efficiency and accuracy by leveraging AI-driven query generation and data retrieval automation.

**Target Users & Use Cases**

**Primary Users:**

* Market Analysts
* Content Strategists
* Channel Managers

**Key Use Cases:**

* Identifying top-performing channels within specific niches.
* Monitoring competitor content strategies – LLM Categorization(Next Development Phase)
* Generating periodic performance reports for internal stakeholders.

**Workflow Pipeline**

1. Generate targeted YouTube search queries via LLM for niche identification.
2. Query YouTube Data API: filter channels based on subscriber thresholds.
3. Resolve and map channel names to unique channel IDs.
4. Retrieve video data (views, upload dates) for selected timeframes (default: 1 month, maximum: 1 year).
5. Conduct asynchronous and batched API calls to optimize response times.
6. Analyze data to determine top-performing channels, average views per video, and upload frequency.
7. Visualize results interactively and enable PDF report generation.

**Core Technologies & Stack**

| **Function** | **Technology Choices** |
| --- | --- |
| LLM & Prompt Orchestration | OpenAI GPT-4, LangChain |
| Backend | Python, FastAPI |
| Frontend | React, TypeScript |
| Visualization Libraries | Plotly, D3.js |
| PDF Generation | Puppeteer or Playwright |

**Optimization Considerations**

* **API Management:** Handle rate limits through batching, asynchronous patterns, and retries.
* **Caching Strategy:** Implement Redis caching for common queries.
* **Pagination:** Employ cursor-based pagination to efficiently manage large datasets.

**Constraints & Risks**

* **LLM Costs:** Potential high operational costs with extensive LLM usage.
* **YouTube API Quotas:** Risk of hitting daily API limits.
* **Latency:** Possible delays in processing large data volumes.
* **PDF Rendering:** Performance bottlenecks in dynamic PDF generation.
* **Compliance:** Adherence to data privacy standards (GDPR/CCPA).

**Expected Outputs & Report Format**

**Graphs & Metrics:**

* Top channels ranked by total views.
* Average views per video.
* Total uploads per channel.

**Deliverables:**

* Interactive analytics dashboard.
* Downloadable PDF reports.

**Stretch Goals / Future Enhancements**

* Scheduled automated reporting (weekly/monthly).
* Slack and email integrations for alerts and notifications.
* Customizable dashboards for different stakeholders.
* Anomaly detection and proactive alerting.
* Specialized Market Analysis.
* Content Tagging – Understanding new popular formats.

**Development Timeline & Milestones**

| **Milestone** | **Start Date** | **End Date** | **Owner** |
| --- | --- | --- | --- |
| Project Kickoff & Requirements | Week 1 | Week 1 | Data & Innovation Team |
| Backend & API Integration | Week 1/2 | Week 1/2 | Data & Innovation Team |
| Frontend Development | Week 1/2 | Week 1/2 | Data & Innovation Team |
| Data Analysis & Visualization | Week 2 | Week 2 | Data & Innovation Team |
| PDF Generation & Optimization | Week 2 | Week 2 | Data & Innovation Team |
| Testing, QA & Feedback | Week 3 | Week 3 | Content/Strategy Teams |
| MVP Launch | TBA | TBA | TBA |
| Full Rollout & Enhancements | TBA | TBA | TBA |

**Success Metrics & Acceptance Criteria**

* Complete workflow automation from query generation to report creation.
* Reduction in manual analysis time.
* User satisfaction rate (via internal survey).
* Stability and professionalism of the analytical dashboard.