HajjSense: Visualizing Pilgrim Experience, Safety, and AR Engagement During Hajj

CSCI 181DV – High-Performance Visualization Project

1. Project Definition & Scope (10 points)

Problem Statement

Hajj and Umrah involve the movement of millions of pilgrims within confined spaces, leading to high crowd density, health risks, and logistical complexity. Managing these factors in real-time is challenging, and traditional systems do not provide adequate insight into pilgrim well-being or crowd dynamics. With the growing role of AR technology, there's an opportunity to visualize these experiences more meaningfully.

Goal: Create a lightweight visualization system to explore pilgrim behavior, fatigue, AR engagement, and incident patterns using simulated data, providing insights for improved safety, navigation, and experience.

Target Users & Use Cases

- **Hajj Organizers:** Monitor crowd conditions, safety alerts, and AR effectiveness
- **Public Health Officers:** Track fatigue, stress, and medical risks in crowded zones
- **AR Developers:** Improve navigation assistance by understanding usage and success rates
- **Researchers:** Analyze behavior patterns and crowd morale across locations

Expected Insights & Outcomes

- Identify high-risk zones for crowding, fatigue, and incidents
- Understand correlations between stress, environmental factors, and safety ratings
- Track AR guidance system adoption and satisfaction
- Visualize emotional and physical state distributions across space and time

Real-World Problems Addressed

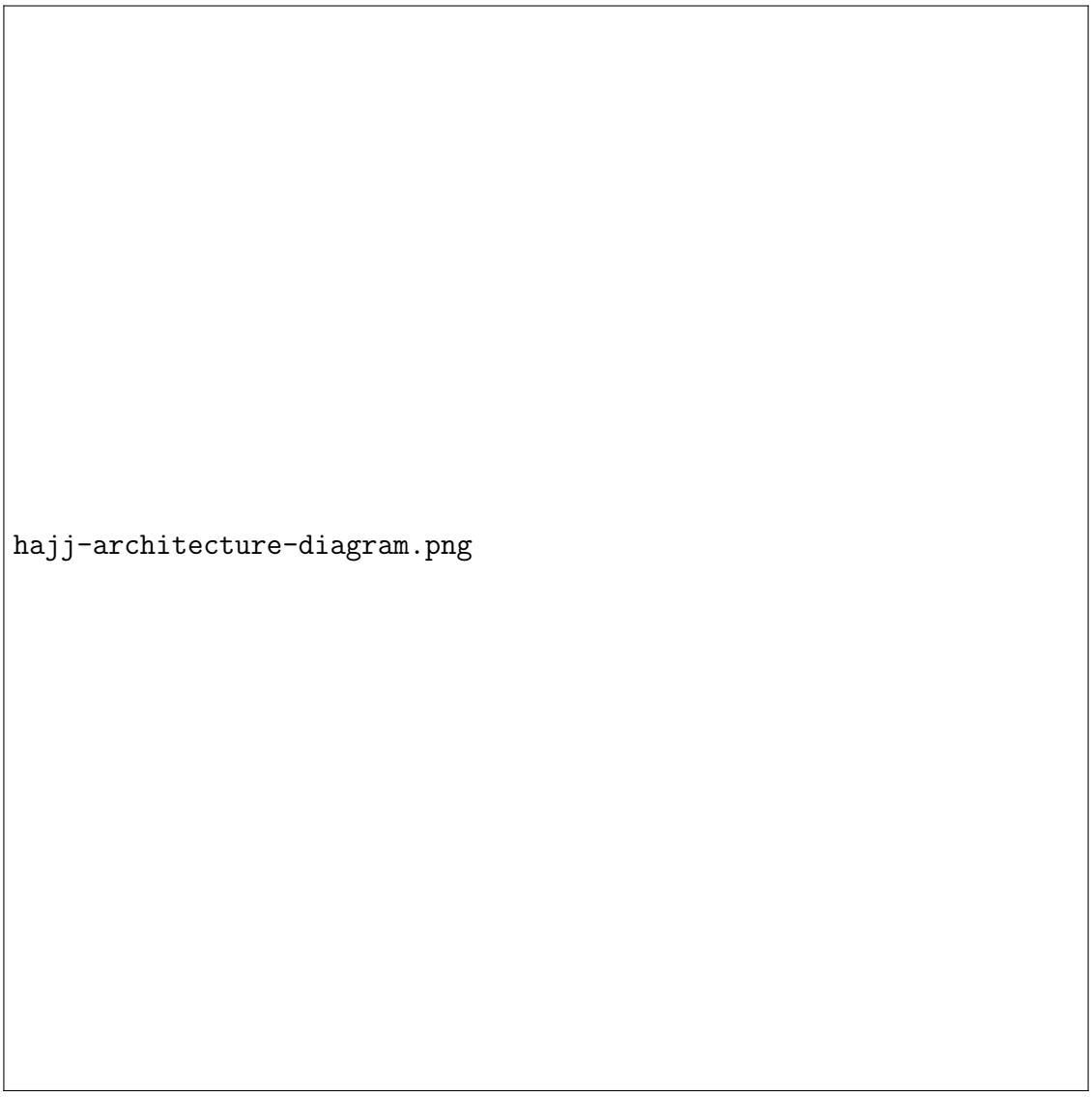
- **Overcrowding at Ritual Sites:** Identify peak congestion areas using heatmaps to prevent stampedes and optimize flow
- **Inefficient Crowd Flow:** Analyze movement speed and spacing to adjust routes and signage
- **Fatigue and Stress Hotspots:** Track fatigue/stress in relation to temperature and density to improve scheduling and rest areas
- **Medical Emergencies in Heat:** Correlate health conditions with environmental factors for early intervention planning
- **Low AR Engagement:** Understand which groups underuse AR navigation and why
- **Emergency Event Mapping:** Visualize incidents to guide emergency team placement
- **Long Wait Times:** Analyze delays at transport and security to reduce queues and improve flow
- **Low Satisfaction by Demographic:** Compare ratings across age and nationality to personalize services
- **Perceived Safety Mismatch:** Match low perceived safety areas to actual risks for targeted improvements

Project Boundaries & Constraints

- Based on simulated dataset from Kaggle (10,000 records)
- No real-time streaming or geolocation integration
- Focused on exploratory and descriptive visualization only
- Implemented entirely with local CSV and in-memory Python tools

2. Technical Design (10 points)

System Architecture



Technology Stack

Layer	Tool/Library	Justification
Backend	Python + Streamlit	Simplifies interaction without server setup
Data Processing	Pandas	Ideal for CSV and tabular simulation data
Storage	CSV files	No database required; easy for testing
Frontend	Altair / Plotly via Streamlit	Interactive visual exploration with minimal setup
Deployment	Localhost / Streamlit Cloud	Easy sharing and running with small files

Data Flow Pipeline

1. Load and clean CSV into memory using Pandas
2. Generate derived features (e.g., crowd zones, AR usage success)
3. Visualize data using Streamlit dashboard widgets and charts

Performance Strategies

- Use `@st.cache` to cache loading and filtered views
- Apply pre-filtering (e.g., by time block or location) to reduce chart load
- Limit visuals to relevant slices (e.g., 1 hour or 1 site at a time)

3. Data Strategy (8 points)

Dataset

Kaggle Dataset: AI Crowd Management Hajj and Umrah (Simulated)
Includes: Location, timestamp, activity, crowd density, stress/fatigue, AR system status, incidents, health condition, satisfaction, and more.

Cleaning & Preprocessing

- Convert timestamps to datetime objects and extract time blocks
- Normalize categorical values (e.g., fatigue, stress levels)
- Derive crowd zones or map coordinates (if visualizing on a map)
- Handle missing or simulated anomalies

Storage Strategy

- Single CSV file loaded with Pandas
- Use filters and slices for subsetting in-memory data

Maintenance Plan

- Versioning new simulations by filename
- Document assumptions for simulations in code and markdown cells
- Extend or update data manually as new formats are tested

4. Visualization Design (7 points)

Mockups (Key Visuals)

- **Crowd Heatmap:** Geographic scatterplot showing `Crowd_Density` across `Location_Lat` and `Location_Long`
- **Fatigue/Stress Over Time:** Line or area chart comparing `Fatigue_Level`, `Stress_Level` by hour or temperature
- **AR Usage Bar Graph:** Compare `AR_System_Interaction` across `Age_Group`, `Activity_Type`, or nationality
- **Incident Frequency Chart:** Horizontal bar chart of `Incident_Type` occurrence by crowd level or AR status
- **Satisfaction**
  Safety Matrix: Scatterplot of `Satisfaction_Rating` vs. `Perceived_Safety_Rating` grouped by experience

Interaction Design

- Dropdown filters for activity type, age group, nationality, AR usage, and incident type
- Sliders for time window and temperature range
- Tooltips with exact values, crowd levels, and incident summaries
- Tabs or collapsible sections for separating health, tech, and satisfaction views

Color & Encoding

Element	Encoding	Rationale
Crowd Density	Red-Green categorical or gradient	Instantly shows risk levels
Fatigue/Stress	Warm tones (yellow-orange-red)	Reflects urgency and discomfort
AR Interaction	Icon or shape encoding	Distinguishes use stages quickly
Incidents	Color by severity/type	Helps triage attention visually

Accessibility Considerations

- Use colorblind-friendly palettes (e.g., ColorBrewer)
- Descriptive tooltips and ARIA labels in charts
- Adjustable font sizes and responsive Streamlit layout
- Icons and patterns as alternatives to color when needed