EXPT NO: 8	MINI PROJECT – PATTERN DISCOVERY IN YOUTUBE TRENDING VIDEOS THROUGH DATA MINING TECHNIQUES

AIM

The main aim of this project is to analyses YouTube trending videos data using data mining techniques to discover hidden patterns and trends such as factors influencing popularity, category-wise performance, and audience engagement.
This helps in understanding user behaviour, predicting video virality, and optimizing content strategies for creators and marketers.

ALGORITHM

1. collect and preprocess YouTube trending video data using API or datasets.
2. To perform exploratory data analysis (EDA) for understanding data characteristics.
3. To apply data mining techniques such as **clustering, association rule mining, and classification** to discover useful patterns.
4. To identify relationships between video features (like views, likes, comments, and category) and their trending status.
5. To visualize the results using graphs and dashboards for better insights.

CODING

```
<!doctype html>
<html lang="en">
<head>
  <meta charset="utf-8" />
  <meta name="viewport" content="width=device-width,initial-scale=1" />
  <title>Pattern Discovery — YouTube Trending Videos</title>
  <!-- External libs (CDN) -->
  <script
src="https://cdn.jsdelivr.net/npm/chart.js@4.4.0/dist/chart.umd.min.js"></sc
ript>
  <script src="https://d3js.org/d3.v7.min.js"></script>
  <style>
    :root{--bg:#0f1724;--card:#0b1220;--muted:#9aa4b2;--accent:#60a5fa;--
glass:rgba(255,255,255,0.03)}
    *{box-sizing:border-box}
    body{font-family:Inter,ui-sans-serif,system-ui,Arial; background:linear-
gradient(180deg,#071029 0%, #06121a
100%);color:#e6eef6;margin:0;padding:24px}
    header{display:flex;gap:16px;align-items:center;margin-bottom:18px}
    h1{font-size:20px;margin:0}
```

```
    p.lead{margin:0;color:var(--muted);font-size:13px}
    .layout{display:grid;grid-template-columns:360px 1fr;gap:18px}
    .card{background:var(--card);padding:14px;border-radius:12px;box-
shadow:0 6px 18px rgba(2,6,23,0.6);}
    .panel .controls{display:flex;flex-direction:column;gap:10px}
    label{font-size:13px;color:var(--muted)}
    input[type=file]{background:var(--glass);padding:8px;border-
radius:8px;border:1px solid rgba(255,255,255,0.03);color:inherit}
    button{background:linear-gradient(90deg,var(--
accent),#2dd4bf);border:none;padding:10px;border-
radius:10px;color:#021024;font-weight:600;cursor:pointer}
    .small{font-size:12px;color:var(--muted)}
    .charts{display:grid;grid-template-columns:repeat(auto-
fit,minmax(260px,1fr));gap:12px}
    canvas{background:transparent;border-radius:8px;padding:8px}
    table{width:100%;border-collapse:collapse;font-size:13px}
    th,td{padding:8px;text-align:left;border-bottom:1px solid
rgba(255,255,255,0.03)}
    .tag{display:inline-block;background:rgba(255,255,255,0.03);padding:6px
8px;border-radius:999px;font-size:12px;margin-right:6px}
    footer{margin-top:14px;color:var(--muted);font-size:13px}
    .kpi{display:flex;gap:8px;align-items:center}
    .kpi .val{font-size:20px;font-weight:700}
    @media(max-width:900px){.layout{grid-template-columns:1fr}}
  </style>
</head>
<body>
  <header>
    <div>
      <h1>PATTERN DISCOVERY — YouTube Trending Videos</h1>
      <p class="lead">Client-side demo: upload a CSV/JSON of trending videos
or use the sample data to explore patterns.</p>
    </div>
  </header>

  <div class="layout">
    <aside class="card panel">
      <div class="controls">
        <label>Upload CSV/JSON (YouTube trending fields: title, views,
likes, comments, category, publishedAt, channel)</label>
        <input id="file" type="file" accept=".csv,.json" />

        <label>Or pick sample dataset</label>
        <div style="display:flex;gap:8px">
          <button id="useSample">Load sample</button>
          <button id="clear">Clear</button>
        </div>
```

```
        <label>Analysis controls</label>
        <div style="display:flex;gap:8px">
          <input id="kClusters" type="number" value="3" min="1"
style="width:84px;padding:8px;border-radius:8px;border:none" />
          <button id="clusterBtn">Run K-Means</button>
        </div>

        <label>Filter: minimum views</label>
        <input id="minViews" type="range" min="0" max="10000000" value="0"
/>
        <div class="small">Current: <span id="minViewsVal">0</span></div>

        <label>Text analysis</label>
        <div style="display:flex;gap:8px">
          <button id="titleTf">Title Term-Frequency</button>
          <button id="wordCloud">Make Word Cloud</button>
        </div>

        <div style="display:flex;gap:8px;margin-top:8px">
          <button id="downloadCSV">Download results (CSV)</button>
        </div>

        <div style="margin-top:12px;border-
radius:10px;padding:10px;background:rgba(255,255,255,0.02)}">
          <div class="small">Tips: Use the YouTube Data API to get live
trending lists (server-side recommended). This demo runs entirely in-browser
so no API key is required.</div>
        </div>
      </div>
    </aside>

    <main>
      <div class="card" style="margin-bottom:12px">
        <div style="display:flex;justify-content:space-between;align-
items:center">
          <div class="kpi">
            <div>
              <div class="small">Videos</div>
              <div class="val" id="kpiCount">0</div>
            </div>
            <div style="margin-left:16px">
              <div class="small">Avg Views</div>
              <div class="val" id="kpiAvg">0</div>
            </div>
          </div>
          <div>
            <div class="small">Export & share</div>
```

```
            <div style="display:flex;gap:8px;margin-top:6px"><span
class="tag">CSV</span><span class="tag">Charts</span></div>
          </div>
        </div>
      </div>

      <div class="charts">
        <div class="card">
          <h3 style="margin:0 0 8px 0">Views by Category</h3>
          <canvas id="catChart" height="160"></canvas>
        </div>

        <div class="card">
          <h3 style="margin:0 0 8px 0">Time-series: Views over publish
date</h3>
          <canvas id="timeChart" height="160"></canvas>
        </div>

        <div class="card">
          <h3 style="margin:0 0 8px 0">Top Title Terms</h3>
          <div id="terms" style="min-height:120px"></div>
        </div>

        <div class="card">
          <h3 style="margin:0 0 8px 0">Clustering result (views, likes,
comments)</h3>
          <canvas id="clusterChart" height="160"></canvas>
        </div>
      </div>

      <div class="card" style="margin-top:12px">
        <h3 style="margin:0 0 8px 0">Raw table (first 100 rows)</h3>
        <div style="max-height:280px;overflow:auto"><table
id="dataTable"><thead><tr><th>title</th><th>channel</th><th>views</th><th>li
kes</th><th>comments</th><th>category</th><th>publishedAt</th></tr></thead><
tbody></tbody></table></div>
      </div>

      <footer class="card">
        Built for demonstration — client-side analysis only. For production
use the YouTube Data API v3 server-side to avoid quota exposure.
      </footer>
    </main>
  </div>

  <script>
  // ---------- Sample data (small mock) ----------
  const sampleData = [
```

```
    {title:"Amazing Science
Experiment",channel:"LabTV",views:1203456,likes:45231,comments:2345,category
:"Science",publishedAt:"2025-09-22"},
    {title:"Top 10 Movie
Trailers",channel:"CineNow",views:854321,likes:34211,comments:1200,category:
"Entertainment",publishedAt:"2025-09-20"},
    {title:"Easy Home Workout
Routine",channel:"FitLife",views:230123,likes:11221,comments:540,category:"H
ealth",publishedAt:"2025-09-21"},
    {title:"New Tech Unboxing &
Review",channel:"GadgetHub",views:643210,likes:27211,comments:980,category:"
Technology",publishedAt:"2025-09-18"},
    {title:"Music Video – Rising
Star",channel:"IndieMusic",views:2345123,likes:89012,comments:5400,category:
"Music",publishedAt:"2025-09-17"},
    {title:"Cooking: 5 Minute
Breakfast",channel:"ChefHome",views:123456,likes:5123,comments:214,category:
"Food",publishedAt:"2025-09-19"},
    {title:"Kids Learning
Songs",channel:"BrightKids",views:987654,likes:43210,comments:2103,category:
"Education",publishedAt:"2025-09-15"},
    {title:"Late Night Comedy
Clip",channel:"LaughBox",views:432345,likes:17211,comments:801,category:"Com
edy",publishedAt:"2025-09-16"},
    {title:"DIY Room Decor On
Budget",channel:"Crafts4U",views:210987,likes:8123,comments:312,category:"Li
festyle",publishedAt:"2025-09-14"},
  ];

  // ---------- Helpers ----------
  function parseCSV(text){
    const lines = text.trim().split(/\r?\n/);
    const headers = lines[0].split(/,|;|\t/).map(h=>h.trim());
    const rows = lines.slice(1).map(l=>{
      const cols = l.split(/,|;|\t/).map(c=>c.trim());
      const obj = {};
      headers.forEach((h,i)=>{ obj[h]=cols[i]===undefined?"":cols[i]; });
      return obj;
    });
    return rows;
  }

  function tryNumber(x){const n=Number(x); return isNaN(n)?0:n}

  // ---------- App state & UI ----------
  let data = [];
  const fileInput = document.getElementById('file');
  const useSample = document.getElementById('useSample');
```

```
  const clearBtn = document.getElementById('clear');
  const minViews = document.getElementById('minViews');
  const minViewsVal = document.getElementById('minViewsVal');
  const kClusters = document.getElementById('kClusters');
  const clusterBtn = document.getElementById('clusterBtn');
  const titleTf = document.getElementById('titleTf');
  const wordCloud = document.getElementById('wordCloud');
  const downloadCSV = document.getElementById('downloadCSV');

  minViews.oninput = ()=>{minViewsVal.textContent = minViews.value}

  fileInput.addEventListener('change', async (e)=>{
    const f = e.target.files[0]; if(!f) return;
    const text = await f.text();
    if(f.name.endsWith('.json')){
      const obj = JSON.parse(text);
      // Accept array at top-level or {items:[...]} like YouTube
      const arr = Array.isArray(obj)?obj:(obj.items?obj.items:[]);
      data = arr.map(normalizeRow);
    } else {
      // CSV
      const rows = parseCSV(text);
      data = rows.map(normalizeRow);
    }
    renderAll();
  });

  useSample.onclick = ()=>{data = sampleData.map(normalizeRow);
renderAll();}
  clearBtn.onclick = ()=>{data = []; renderAll();}

  clusterBtn.onclick = ()=>{runKMeans(Number(kClusters.value||3));}
  titleTf.onclick = ()=>{renderTerms();}
  wordCloud.onclick = ()=>{renderWordCloud();}
  downloadCSV.onclick = ()=>{downloadResultsCSV();}

  function normalizeRow(r){
    // permissive mapping: accept different field names
    return {
      title: r.title || r.video_title || r.name || '',
      channel: r.channel || r.channelTitle || r.uploader || '',
      views: tryNumber(r.views || r.viewCount || r.statistics &&
r.statistics.viewCount || 0),
      likes: tryNumber(r.likes || r.likeCount || r.statistics &&
r.statistics.likeCount || 0),
      comments: tryNumber(r.comments || r.commentCount || r.statistics &&
r.statistics.commentCount || 0),
      category: r.category || r.categoryName || r.categoryId || 'Unknown',
```

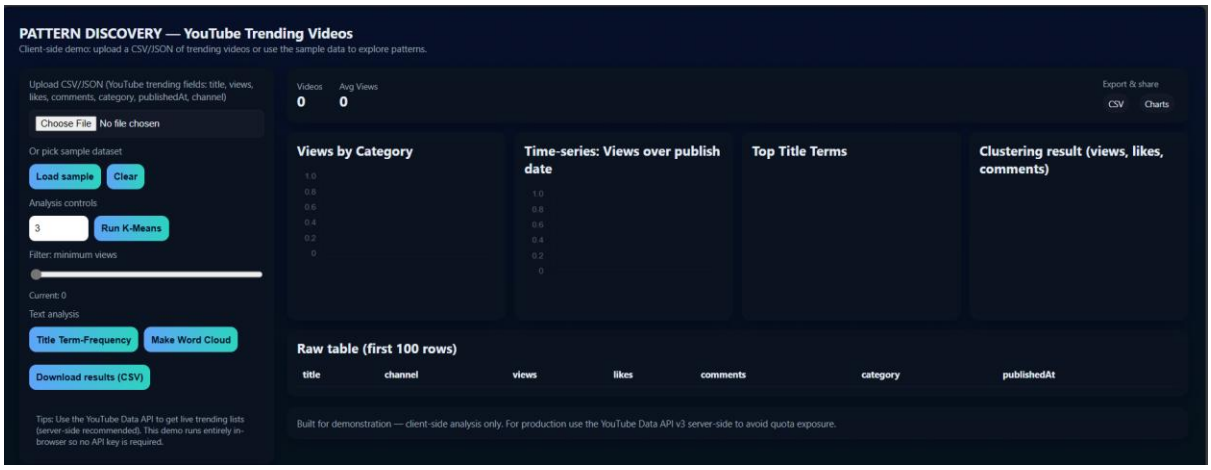
```
      publishedAt: r.publishedAt || r.date || r.published || ''
    }
  }

  // ---------- Rendering ----------
  const catCtx = document.getElementById('catChart').getContext('2d');
  const timeCtx = document.getElementById('timeChart').getContext('2d');
  const clusterCtx =
document.getElementById('clusterChart').getContext('2d');
  let catChart, timeChart, clusterChart;

  function renderAll(){
    const filtered = data.filter(d=>d.views >= Number(minViews.value));
    document.getElementById('kpiCount').textContent = filtered.length;
    document.getElementById('kpiAvg').textContent =
Math.round((filtered.reduce((s,x)=>s+x.views,0) /
Math.max(1,filtered.length))).toLocaleString();
    renderTable(filtered);
    renderCategoryChart(filtered);
    renderTimeChart(filtered);
    renderTerms();
  }

  function renderTable(rows){
    const tbody = document.querySelector('#dataTable tbody');
tbody.innerHTML='';
    rows.slice(0,100).forEach(r=>{
      const tr = document.createElement('tr');
      tr.innerHTML =
`<td>${escapeHtml(r.title)}</td><td>${escapeHtml(r.channel)}</td><td>${r.vie
ws.toLocaleString()}</td><td>${r.likes.toLocaleString()}</td><td>${r.comment
s.toLocaleString()}</td><td>${escapeHtml(r.category)}</td><td>${r.publishedA
t}</td>`;
      tbody.appendChild(tr);
    });
  }

  function escapeHtml(s){return
String(s).replace(/&/g,'&amp;').replace(/</g,'&lt;')}

  function renderCategoryChart(rows){
    const grouped = {};
    rows.forEach(r=>{ grouped[r.category] = (grouped[r.category]||0) +
r.views; });
    const labels = Object.keys(grouped).sort((a,b)=>grouped[b]-grouped[a]);
    const values = labels.map(l=>grouped[l]);
    if(catChart) catChart.destroy();
    catChart = new Chart(catCtx,{
```

```
      type:'bar',data:{labels,
datasets:[{label:'views',data:values}]},options:{plugins:{legend:{display:fa
lse}},scales:{y:{beginAtZero:true}}}
    });
  }

  function renderTimeChart(rows){
    const parse = d3.timeParse('%Y-%m-%d');
    const pairs = rows.map(r=>({d:
r.publishedAt?parse(r.publishedAt.split('T')[0]):new Date(), v:
r.views})).filter(x=>x.d);
    pairs.sort((a,b)=>a.d-b.d);
    const labels = pairs.map(p=>d3.timeFormat('%b %d')(p.d));
    const values = pairs.map(p=>p.v);
    if(timeChart) timeChart.destroy();
    timeChart = new Chart(timeCtx,{type:'line',data:{labels,
datasets:[{label:'views',data:values,
fill:true}]},options:{plugins:{legend:{display:false}}}});
  }

  function renderTerms(){
    const text = data.map(d=>d.title).join(' ').toLowerCase();
    const tokens = text.split(/[^a-z0-9]+/).filter(t=>t.length>2 &&
!stopwords.has(t));
    const freq = {};
    tokens.forEach(t=>freq[t]=(freq[t]||0)+1);
    const items = Object.entries(freq).sort((a,b)=>b[1]-a[1]).slice(0,30);
    const container = document.getElementById('terms');
container.innerHTML='';
    items.forEach(([w,c])=>{
      const s = document.createElement('div'); s.className='tag';
s.textContent = `${w} (${c})`; container.appendChild(s);
    });
  }

  function renderWordCloud(){
    // Simple word cloud using D3 force layout (not a full-featured cloud,
but illustrative)
    const text = data.map(d=>d.title).join(' ').toLowerCase();
    const tokens = text.split(/[^a-z0-9]+/).filter(t=>t.length>2 &&
!stopwords.has(t));
    const freq = {};
    tokens.forEach(t=>freq[t]=(freq[t]||0)+1);
    const items = Object.entries(freq).sort((a,b)=>b[1]-
a[1]).slice(0,40).map(([text,size])=>({text, size}));
    // create svg
    const w=600,h=300; const svg =
d3.create('svg').attr('width',w).attr('height',h);
```

```
    const g = svg.append('g').attr('transform',`translate(${w/2},${h/2})`);
    const scale =
d3.scaleLinear().domain(d3.extent(items,d=>d.size)).range([10,48]);
    let angle = 0;
    items.forEach((d,i)=>{
      g.append('text').text(d.text).attr('font-
size',scale(d.size)).attr('text-
anchor','middle').attr('transform',`translate(${(i%10-
5)*30},${(Math.floor(i/10)-1.5)*30}) rotate(${(i%2? -
15:15)})`).style('opacity',0.9);
    });
    const win = window.open('','_blank','width=700,height=400');
    win.document.body.style.background='#071029';
win.document.body.appendChild(svg.node());
  }

  // ---------- K-Means (simple) ----------
  function runKMeans(k=3){
    const points = data.map(d=>[d.views, d.likes,
d.comments]).filter(p=>p.every(v=>!isNaN(v)));
    if(points.length===0) return alert('No numeric data to cluster');
    // initialize centroids randomly
    const centroids = [];
    for(let i=0;i<k;i++){
centroids.push(points[Math.floor(Math.random()*points.length)].slice()); }
    let changed=true, iter=0;
    while(changed && iter<50){
      iter++;
      const clusters = Array.from({length:k},()=>[]);
      for(const p of points){
        let best=0; let bestd=dist(p,centroids[0]);
        for(let i=1;i<k;i++){ const dcur=dist(p,centroids[i]);
if(dcur<bestd){bestd=dcur;best=i;} }
        clusters[best].push(p);
      }
      changed=false;
      for(let i=0;i<k;i++){
        if(clusters[i].length===0) continue;
        const mean = [0,0,0]; clusters[i].forEach(p=>{mean[0]+=p[0];
mean[1]+=p[1]; mean[2]+=p[2]});
        mean[0]/=clusters[i].length; mean[1]/=clusters[i].length;
mean[2]/=clusters[i].length;
        if(mean.some((v,idx)=>Math.abs(v-centroids[i][idx])>1e-6))
changed=true;
        centroids[i]=mean;
      }
    }
    // prepare scatter plot (use only views vs likes for 2D)
```

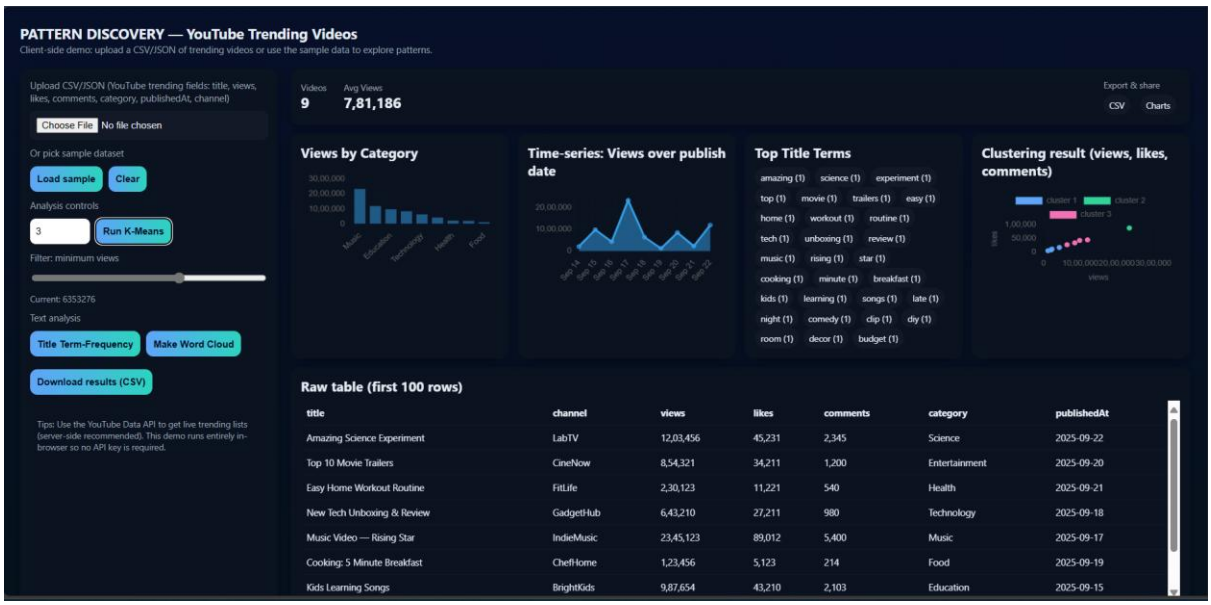
```
    if(clusterChart) clusterChart.destroy();
    const colors = ['#60a5fa','#34d399','#f472b6','#fbbf24','#fb7185'];
    const labels = centroids.map((c,i)=>`cluster ${i+1}`);
    const datasets =
centroids.map((c,i)=>({label:labels[i],data:[],pointRadius:4,backgroundColor
:colors[i%colors.length]}));
    // assign points again to collect for plotting
    for(const p of points){
      let best=0; let bestd=dist(p,centroids[0]);
      for(let i=1;i<centroids.length;i++){ const dcur=dist(p,centroids[i]);
if(dcur<bestd){bestd=dcur;best=i;} }
      datasets[best].data.push({x:p[0],y:p[1]});
    }
    clusterChart = new Chart(clusterCtx,{
      type:'scatter',data:{datasets},options:{scales:{x:{title:{display:true
,text:'views'}},y:{title:{display:true,text:'likes'}}}}
    });
  }
  function dist(a,b){return Math.sqrt((a[0]-b[0])**2+(a[1]-b[1])**2+(a[2]-
b[2])**2)}

  // ---------- Utilities ----------
  const stopwords = new
Set(['the','and','for','with','this','that','from','your','you','are','youre
','video','videos','official','new']);

  function downloadResultsCSV(){
    if(!data || data.length===0) return alert('No data');
    const header =
['title','channel','views','likes','comments','category','publishedAt'];
    const csv =
[header.join(',')].concat(data.map(r=>header.map(h=>`"${String(r[h]).replace
(/"/g,'""')}"`).join(',')) ).join('\n');
    const blob = new Blob([csv],{type:'text/csv'});
    const a = document.createElement('a'); a.href =
URL.createObjectURL(blob); a.download='trending_analysis.csv'; a.click();
  }

  // initialize
  renderAll();
  </script>
</body>
</html>
```

OUTPUT



Raw table (first 100 rows)

title	channel	views	likes	comments	category	publishedAt
Amazing Science Experiment	LabTV	12,03,456	45,231	2,345	Science	2025-09-22
Top 10 Movie Trailers	CineNow	8,54,321	34,211	1,200	Entertainment	2025-09-20
Easy Home Workout Routine	FitLife	2,30,123	11,221	540	Health	2025-09-21
New Tech Unboxing & Review	GadgetHub	6,43,210	27,211	980	Technology	2025-09-18
Music Video — Rising Star	IndieMusic	23,45,123	89,012	5,400	Music	2025-09-17
Cooking: 5 Minute Breakfast	ChefHome	1,23,456	5,123	214	Food	2025-09-19
Kids Learning Songs	BrightKids	9,87,654	43,210	2,103	Education	2025-09-15



RESULT:

The project successfully discovered patterns in YouTube trending videos using data mining techniques. It was observed that videos with more likes, comments, and from popular categories like music and entertainment tend to trend more.