DATA 400 Mini-Project Idea 2

Title: *Comparing Bank Account Fees & Features Across Retail Banks*

1. Research Question

How do monthly fees, minimum balance requirements, and ATM charges vary across major U.S. banks, and which banks are more consumer-friendly for different customer types (e.g., students, low-balance users, frequent ATM users)?

2. Motivation

Bank account fees affect millions of people. Even small differences in monthly maintenance charges or ATM fees add up quickly. By scraping and comparing fee schedules from multiple banks, this project highlights transparency in banking and identifies which institutions are more favorable for common consumer profiles.

This project is simple but practical: it turns unstructured public information (bank product pages) into structured data that can be analyzed and visualized.

3. Data Collection
- **Source:** Public bank websites (checking/savings account product pages, fee schedule tables, promotional account information).
- **Method:** Web scraping using requests, BeautifulSoup, and pandas.read_html.
- **Scope:** 6 - 8 major U.S. banks (e.g., Bank of America, Chase, Wells Fargo, PNC, Citi, Capital One, Ally, Discover).

4. Variables to capture:
- Monthly maintenance fee.
- Minimum balance requirement.
- ATM out-of-network fee.
- Overdraft fee.
- Savings APY (if shown).
- Promotional offers (if applicable)

5. Planned Analysis

1. **Exploratory Data Analysis (required):**
   - Compare average monthly fees, ATM fees, and overdraft charges across banks.
   - Visualize distributions using bar charts and boxplots.
   - Rank banks by "friendliness" for specific profiles:
     - *Student/low balance:* prefers no minimum balance, no monthly fee.
     - *Frequent ATM user:* cares about ATM fees.
2. **Optional Model:**
   - Simple clustering of banks based on fee structures and APYs to group "consumer-friendly" vs. "fee-heavy" banks.
3. **Interpretation:**
   - Identify which features differ most between banks.
   - Discuss transparency issues: some banks post fees clearly, others bury them.

6. Ethical Considerations

- Only scrape publicly available product/fee pages.
- Respect each site's robots.txt file and throttle requests.
- Credit sources clearly in README.
- No collection of personal or customer account data.

7. Expected Deliverables

- Tidy dataset (CSV) with bank fee data.
- 3 - 4 clear plots comparing fees and features.
- 3 - 5 bullet-point insights (e.g., "online banks cluster with no monthly fees, but traditional banks charge $10 - 15/month unless balance ≥ $1,500").
- 6 - 8 presentation slides summarizing the question, scraping process, visuals, and takeaways.