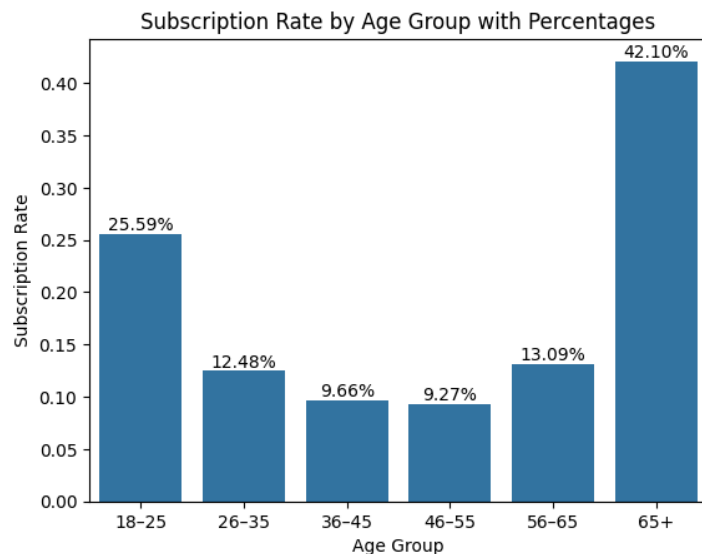# Feb 25, 2026 | Team Project

Notes -
- Business Problem
- Business Motivation
- Intended Audience
- Expected Impact
- Compiled Questions:
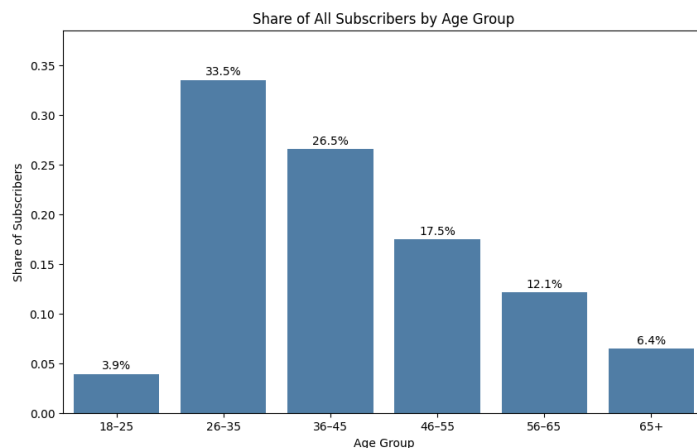   a. **01) Demographics** **kerensa.ca@gmail.com**
        i. Which age groups are most likely to subscribe to a term deposit? (Jarsmeka)

Subscription Rate by Age Group with Percentages

(Likelihood within each age group ← finding which groups more likely to subscribe)

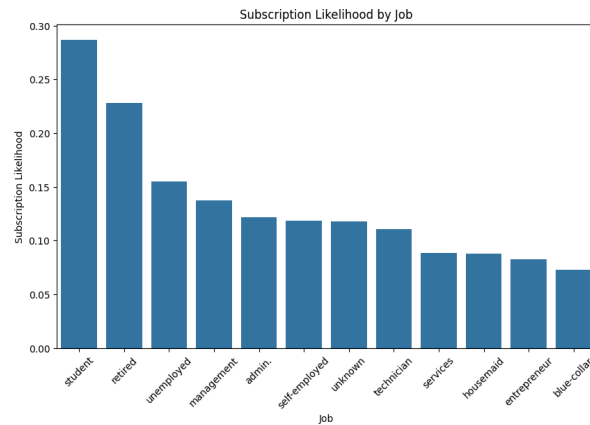### Share of All Subscribers by Age Group

(Volume - 26-45age = most subscriber coz they are the largest group, but their subscription rate is low (not responsive) ← above graph)
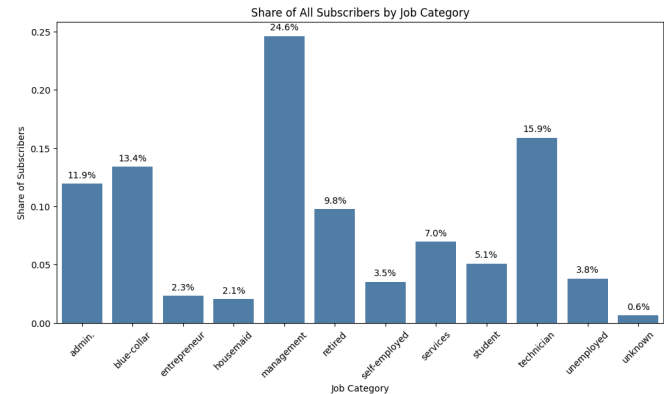
** 65+ → high conversion rate/potential; 18-25 → responsive; 26-45: large group but low responsiveness)

        ii. Are retired customers more likely to subscribe than working professionals? (Vince)
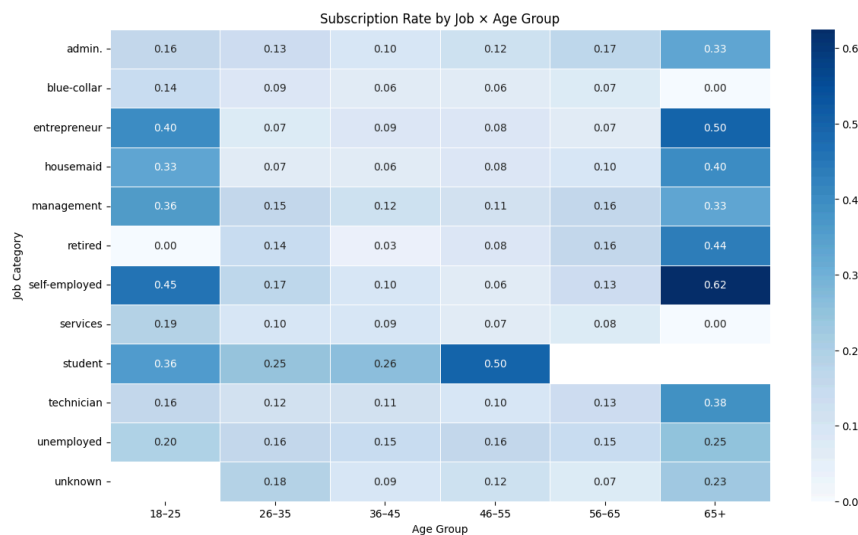
Likelihood:

Volume:

- Student and retired are easiest to convert, but they are small groups
    - Great to target, high ROI campaign
- Volume mostly come from working professionals (management, technician, blue-collar, admin)
    - Low likelihood, but large group so even small improvement in conversion can produce big gains
- High likelihood ≠ high volume (same as age group result)

iii.   Does a customer's job influence subscription probability? (Vince, Jarsmeka)

Age vs job analysis

Heatmap (subscription likelihood for every job-age combination)

- self-employed(65+) highest subscription rate among all, retired (65+) also high
- student (18-45): high likelihood in both young and mid-age group
- blue-collar: low across all age group ← job effect dominate age effect
- management: young age and 56+ has moderate likelihood
- Entrepreneur & Housemaid: relatively high likelihood (18-25 & 65+)

** some job (incl. Student, retired, self-employed) has high propensity
Blue-collar (in all age) & service: low propensity **

Grouped Barplot (Volume)

Subscription Rate by Job and Age Group
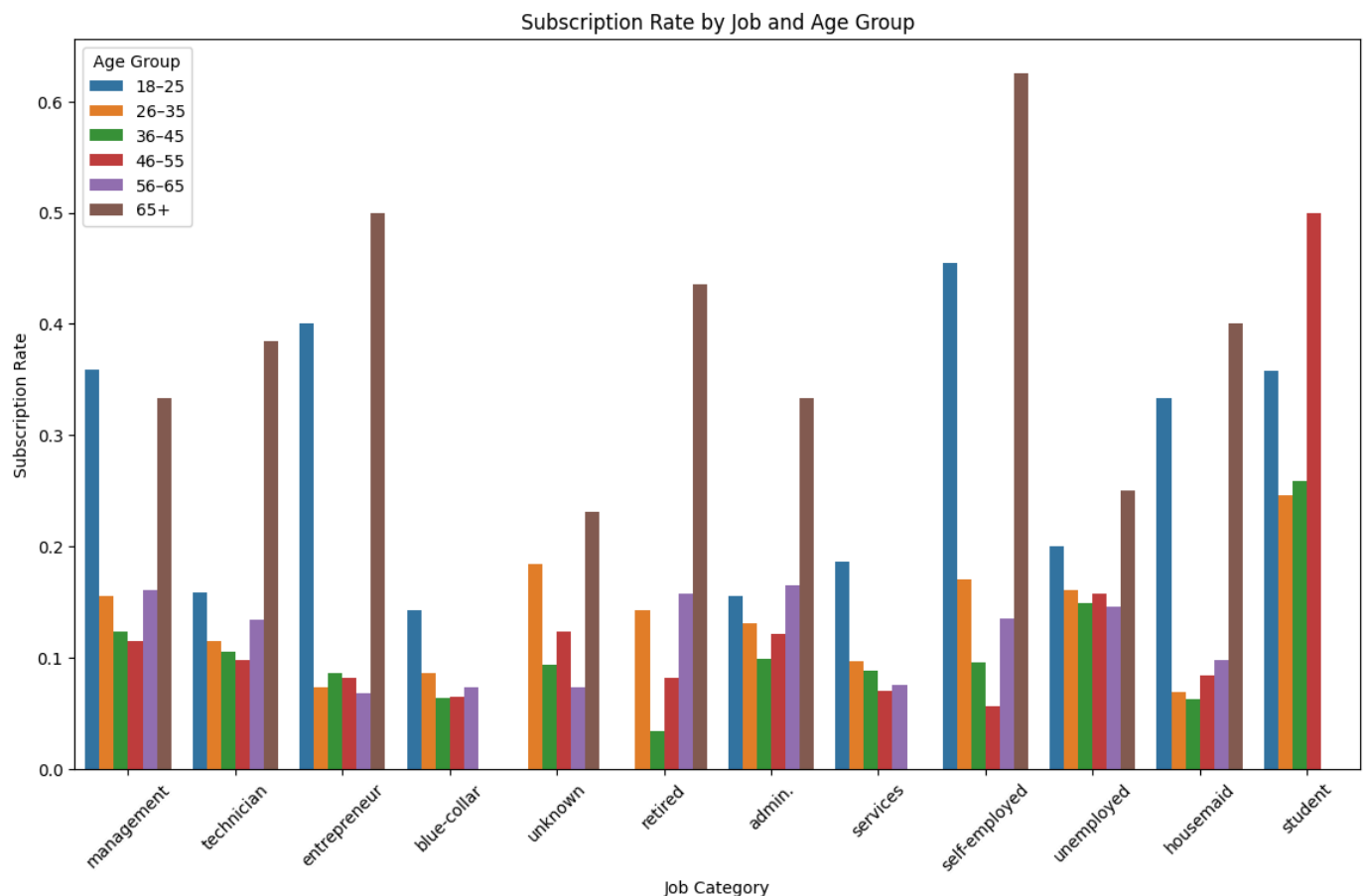
iv.   Does a customer's education level impact subscription? (Jarsmeka)
v.   How does marital status impact subscription behavior? (Jarsmeka)

b. 02) Financial Profile J M

i.   Does higher account balance increase likelihood of subscription? (Vince, Jarsmeka)
ii.   Does having a housing loan affect subscription probability? (Vince)
iii.   Does having a personal loan affect subscription behavior? (Vince)
iv.   Does having credit in default reduce subscription probability? (Vince)
v.   Is there an interaction between balance and loan status? (Vince)
vi.   Can subscription likelihood be linked to other products (e.g., loans)? (Kerensa)
vii.   How many loans are personal vs housing loans? (Jarsmeka)

c. 03) Campaign Strategy phuc.nguyen.ufo@gmail.com

i.   Does contact type (cellular vs telephone) affect subscription success? (Jarsmeka)
ii.   Is there an optimal number of contacts in a campaign? (Vince)
iii.   Does subscription probability decrease after too many calls? (Vince)
iv.   What months have the highest vs lowest subscription success? (Jarsmeka, Vince)
v.   Are certain days of the week more effective? (Vince)

      vi.   Is there evidence of seasonal patterns in subscription behavior? (Vince)

      vii.   How does call duration relate to subscription? (Vince)

d. 04) Previous Campaign Effectiveness **J M**

      i.   How does the outcome of a previous campaign affect current success? (Vince, Jarsmeka)

      ii.   Are customers with previous successful campaigns more likely to subscribe again? (Vince)

      iii.   Are customers previously unsuccessful unlikely to convert again? (Vince)

      iv.   Does time since last contact (pdays) influence success probability? (Vince)

      v.   Are existing customers more likely to subscribe than new ones? (Kerensa)

e. 05) Segmentation & Predictions phuc.nguyen.ufo@gmail.com

      i.   What combination of age, job, balance, and loan status yields the highest conversion rate? (Vince)

      ii.   Can we identify low-probability segments to avoid targeting? (Vince)

      iii.   Can customers be ranked by likelihood of subscription? (Vince)

      iv.   Does combining variables provide stronger predictive power than single variables? (Kerensa)

      v.   Does the combination of age + job + loan status improve prediction compared to single factors? (Kerensa)

- Risks and Unknowns

a. Risks kerensa.ca@gmail.com

      i.   The dataset may be heavily skewed toward "no" subscriptions compared to "yes" (class imbalance). (Vince, Jarsmeka)

      ii.   The dataset only includes campaigns from a Portuguese bank, which may limit applicability in other countries (e.g., Canada). (Kerensa, Jarsmeka)

      iii.   The data is older than 2013, so customer behavior and economic conditions may have changed (e.g., financial crises, COVID), reducing predictive relevance today. (Jarsmeka, Kerensa)

      iv.   Class imbalance may affect model accuracy and interpretation of results. (Vince)

      v.   Data quality concerns: "unknown" values in variables such as job, education, contact, and poutcome. (Kerensa)

      vi.   The "other" category in poutcome lacks explanation and may reduce interpretability. (Kerensa)

      vii.   Extreme outliers (e.g., very high balances or very long call durations) may skew results. (Kerensa)

      viii. Predictive models may perform well on historical data but fail in real-world future campaigns. (Kerensa)

b. Unknowns

      i.   The cost of running each marketing campaign. (Jarsmeka)

      ii.   The profit generated per successful subscription. (Jarsmeka)

      iii.   Why customers choose not to subscribe (behavioral or psychological reasoning). (Jarsmeka)
           How long customers maintain their term deposit after subscribing. (Jarsmeka)

      iv.   Whether loans are actively being used or simply held. (Jarsmeka)

    v.   Other non-demographic factors influencing subscription (e.g., personal motivations). (Jarsmeka)
    vi.   Whether the campaign strategy or call scripts changed over time. (Jarsmeka)
    vii.   External economic factors that may have influenced customer decisions. (Vince)
    viii.   Important variables not included in the dataset (e.g., income, digital engagement, geographic location). (Kerensa)
    ix.   Time of day of calls (morning vs evening). (Kerensa)
    x.   Whether customers consented to repeated contact (ethical consideration). (Kerensa)
    xi.   Correlation does not imply causation — observed relationships may not represent true drivers. (Kerensa)

Action items
- [ ] Analysis and Visualization to be completed by Friday
  - [ ] Initial Analysis, Demographics and Project Proposal - Kerensa
  - [ ] Financial Profile and Previous Campaign Effectiveness, Initial Showcase Slideshow - Jarsmeka
  - [ ] Campaign Effectiveness and Segmentation/Prediction - Vince