R1: Group 22 9/12 → 4.5/6

1. Formatting: 0.75/0.75

all margins 2.5cm

12 pt size

no raw R code or output

max **7** pages

informative title

member names on all pgs

all pages numbered

no blurry plots (**NOT png**)

2. Introduction/Background: 1/1

brief statement of scientific question

all variables defined

3. EDA: 1.5/2 figures too small

univariate numerical

univariate graphical

bivariate numerical (cor)

bivariate graphical

4. Model fitting: 1.25/2

state how model fitted (ie, LS) First write the model mathematically

CLEARLY describe how model selected - AIC considers p-values?

define all terms VIF

5. Model assessment: 1.25/2

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):
qq normal plot of residuals, - square
residuals vs. fitted

- clearly interpret plots

5.75/7.75

1/1 6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

0.75/1.25 7. Plots:

label size (not too small) (captions) not completely informative

placement **NOT BLURRY**

0.5/1 8. Conclusions somewhat generic

recap analysis (state main findings)

1/1 9. Language quality:

poor satisfactory (good) excellent

10. Other comments:

- no refs

3.25/4.25

R1: Group 3 | 7.25/12 → 3.625/6

1. Formatting:

0.5/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output** → no R formulas in plots
- max **7** pages
- informative title
- member names on all pgs
- all pages numbered
- no blurry plots (**NOT png**)

2. Introduction/Background: 1/1

- brief statement of scientific question
- all variables defined

3. EDA: 0.75/2

- univariate numerical — Don't need boxplot Figure1 — SQUARE QQ
- univariate graphical → all vars
- bivariate numerical (cor)
- bivariate graphical — SQUARE Pairs plots

4. Model fitting: 1/2

- First write the model mathematically
- state how model fitted (ie, LS)
- CLEARLY describe how model selected
- define all terms — R^2/R^2 adj — where are model df? only 1 more df, why not use that model?

5. Model assessment: 1/2

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):

- qq normal plot of residuals, — SQUARE
- residuals vs. fitted

indep obs ≡ indep errors ⇒ redundant

→ clearly explain plot interp

4.25/7.75

1/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

0.75/1

7. Plots:

1.25/1

label size (not too small) captions

placement **NOT BLURRY**

0.25/1

8. Conclusions

recap analysis very generic state main findings

1/1

9. Language quality:

poor satisfactory good excellent

10. Other comments:

3/4.25

R1: Group 32 7/12 → 3.5/6

1. Formatting: 0.75/0.75

all margins 2.5cm

12 pt size

no raw R code or output

max **7** pages

informative title

member names on all pgs

all pages numbered

no blurry plots (**NOT png**)

2. Introduction/Background: 1/1

brief statement of scientific question

all variables defined (Hard to read in line – put in 'quotation'-type paragraph

3. EDA: 0.5/2

univariate numerical

univariate graphical

bivariate numerical (cor)

bivariate graphical

↳ histograms not boxplots – all vars

4. Model fitting: 0.75/2

First write model mathematically

state how model fitted (ie, LS)

CLEARLY describe how model selected How does backward achieve 'most accurate' model?? What does this mean?

define all terms AIC

P-Value criterion?

5. Model assessment: 1.25/2

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):

qq normal plot of residuals,

residuals vs. fitted

careful + correct plot interp

4.25

0.5/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

1/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.25/1

8. Conclusions

very vague/generic

recap analysis state main findings

1/1

9. Language quality:

poor satisfactory good excellent

10. Other comments:

2.75/4.25

R1: Group 33 7.75/12 → 3.875/6

1. Formatting: 0.75/0.75

all margins 2.5cm

12 pt size

no raw R code or output

max **7** pages

informative title

member names on all pgs

all pages numbered

no blurry plots (**NOT png**)

1/1 2. Introduction/Background:

brief statement of scientific question

all variables defined

1/2 3. EDA: SQUARE QQ

univariate numerical

bivariate numerical (cor)

univariate graphical

↳ histograms (all vars) not boxplots

bivariate graphical

1/2 4. Model fitting:

first write model mathematically

state how model fitted (ie, LS)

CLEARLY describe how model selected

define all terms stepwise/AIC

5. Model assessment:

1/2

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):
qq normal plot of residuals,
residuals vs. fitted

- you don't 'evaluate validity'
- more careful plot interpretation

4.75/7.75

1/1 6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

0.5/1.25 7. Plots: informative 'pretty' labels

label size (not too small) captions

placement **NOT BLURRY**

0.5/1 8. Conclusions

recap analysis state main findings

→ use paragraphing (more)

1/1 9. Language quality:

poor satisfactory good excellent

10. Other comments:

- no refs

3/4.25

R1: Group 34 6.75/12 → 3.375/6

1. Formatting:

0.5/0.75

all margins 2.5cm

12 pt size

no raw R code or output
↳ R formulas in plots + tables

max **7** pages

informative title

member names on all pgs

all pages numbered

no blurry plots (**NOT png**)

2. Introduction/Background:

1/1

brief statement of scientific question

all variables defined

3. EDA:

1/2

univariate numerical

univariate graphical

bivariate numerical (cor) (some hard to read)

bivariate graphical square
↳ + all pairs

4. Model fitting:

1/2

First write model mathematically

state how model fitted (ie, LS)

CLEARLY describe how model selected

define all terms

⇒ Don't need model R^2/R^2_{adj} to tell you direction of association, you can get that from correlation

5. Model assessment:

0.75/1

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):
qq normal plot of residuals, – line is a reference line (not 'theoretical' line)
residuals vs. fitted

← incorrect interpretation
⇒ carefully interpret plots

4.25/7.75

0.5/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

0.75/1.25

7. Plots: -no R formulas

label size (not too small) captions

placement **NOT BLURRY**

0.25/1

8. Conclusions

recap analysis too short/vague state main findings

1/1

9. Language quality:

poor satisfactory good excellent

10. Other comments:

-Table 2 p-values >1

-You usually don't need 2 highly corr vars in model → multicollinearity

2.5/4.25

R1: Group 35 9.5/12 → 4.75/6

1. Formatting:

0.75/0.75

- all margins 2.5cm
- 12 pt size — refs
- **no raw R code or output**
- max **7** pages
- informative title
- member names on all pgs
- all pages numbered
- no blurry plots (**NOT png**)

1/1

2. Introduction/Background:

- brief statement of scientific question
- all variables defined — hard to read inline, put in separate paragraph

1/2

3. EDA: var 1: ~~~ etc
var 2: ~~~

- univariate numerical
- bivariate numerical (cor) matrix
- univariate graphical
- bivariate graphical — all pairs

0.75/2

4. Model fitting: First write the mathematical model

- state how model fitted (ie, LS)
- CLEARLY describe how model selected
- define all terms AIC / stepwise (Forward/Backward)

1.5/2

5. Model assessment:

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):
- qq normal plot of residuals, — SQUARE
- residuals vs. fitted

- assumption 'satisfied' (not 'validated')
- Carefully explain interpretation (don't use the word 'significant')

5/7.75

1/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

1.25/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

1/1

8. Conclusions

recap analysis state main findings

1.25/1

9. Language quality:

poor satisfactory good excellent

10. Other comments:

4.5/4.25

R1: Group 3b 5.25/12 → 2.625/6

1. Formatting:

0.75/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **7** pages
- informative title
- member names on all pgs
- all pages numbered
- no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

- brief statement of scientific question
- all variables defined → in intro, not in separate sections → very spread out so hard to read/keep track

0.75/2

3. EDA:

- univariate numerical → all in 1 table
- bivariate numerical (cor)
- univariate graphical
- bivariate graphical all pairs

0.75/2

4. Model fitting: – First write model mathematically
- state how model fitted (ie, LS) – CLEARLY, in fitting section
- CLEARLY describe how model selected
- define all terms vif / stepwise

0/2

5. Model assessment: not done

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):
qq normal plot of residuals,
residuals vs. fitted

3/7.75

0.5/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

0.75/1.25

7. Plots:

~~label size~~ (not too small) captions

placement **NOT BLURRY**

0/1

8. Conclusions not done

recap analysis state main findings

1/1

9. Language quality:

poor satisfactory good excellent

10. Other comments:

- no refs

- incomplete

2.25/4.25

R1: Group 37 7/12 → 3.5/6

1. Formatting:

0.5/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output** (R warning + R formulas)
- max **7** pages
- informative title
- member names on all pgs
- all pages numbered
- no blurry plots (**NOT png**) – too many digits

2. Introduction/Background: 4/1

- brief statement of scientific question
- all variables defined

3. EDA: 0.5/2

- univariate numerical
- univariate graphical
- bivariate numerical (cor)
- bivariate graphical all pairs

4. Model fitting: 1/2 first write model mathematically

- state how model fitted (ie, LS)
- CLEARLY describe how model selected (also could consider AIC/backward/R^2_{adj}/forward)
- define all terms

↳ more clearly, you are doing it iterative NOT 'continuously'

5. Model assessment: 0.75/2

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):
qq normal plot of residuals, – SQUARE
residuals vs. fitted

assumptions 'correct' ?? more like 'approximately satisfied'

3.75/7.75 – incorrect plot interps, not all interpretations given

0.5/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

1.25/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.5/1

8. Conclusions

(*) interpretation

recap analysis incomplete state main findings

1/1

9. Language quality:

poor satisfactory (good) excellent

10. Other comments:

- no refs (put at end, not in text)
- Don't need table S1
- selection not 'propagation'
- highlight lowest AIC in tables

(*) cannot determine causation, only association

3.25/4.25

R1: Group 39 9.25/12 → 4.625/6

1. Formatting:

0.75/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **7** pages
- informative title
- member names on all pgs
- all pages numbered
- no blurry plots (**NOT png**)

1/1

2. Introduction/Background:

- brief statement of scientific question
- all variables defined

3. EDA: – put Table 1 BELOW the description

1/2

- univariate numerical
- bivariate numerical (cor)
- univariate graphical
- bivariate graphical

– Histograms not boxplots ↳ all pairwise scatter plots

1.5/2

4. Model fitting:

- ✓ state how model fitted (ie, LS)
- CLEARLY describe how model selected Explain Likelihood method
- define all terms

5. Model assessment:

1/2

- CLEARLY state model assumptions:
  1. errors have mean 0
  2. errors are homoscedastic (same variance)
  3. errors are uncorrelated
  4. errors are normally distributed

- carry out assessment (graphics):
  qq normal plot of residuals,
  residuals vs. fitted

– careful interpretation of plots

5.25/7.75

1/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

1.25/1.25

7. ~~Plots~~:

label size (not too small) captions

placement **NOT BLURRY**

0.75/1

8. Conclusions (+EDA) recap analysis (*)interpretation state main findings

1/1

9. Language quality:

poor satisfactory (good) excellent

10. Other comments:

(*) Your 'insights' are all framed as 'suggestions, but this is not correct - predicted mortgage yield is (positive when all predictors 0 & etc) - also, you don't need the reg coef to determine correlation, just compute corr.

4/4.25

R1: Group 40 9/12 → 4.5/6

1. Formatting:

0.75/0.75 all margins 2.5cm

12 pt size

no raw R code or output

max **7** pages

informative title

member names on all pgs

all pages numbered

no blurry plots (**NOT png**)

-too many digits

1/1 2. Introduction/Background:

brief statement of scientific question

all variables defined (response? - make more clear)

2/2 3. EDA: -Don't need Figures 2/3

univariate numerical

univariate graphical

bivariate numerical (cor)

bivariate graphical

1.5/2 4. Model fitting:

state how model fitted (ie, LS)

CLEARLY describe how model selected

define all terms

5. Model assessment:

0.75/2 CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):
qq normal plot of residuals,
residuals vs. fitted

-plot interpretations not clear

6/7.75

0.5/1 6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

1/1.25 7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.5/1 8. Conclusions use paragraphs (*) interpretation

(+ EDA recap analysis state main findings

1/1 9. Language quality:

poor satisfactory good excellent

10. Other comments:

- Schaaf (not SHAAF)

- model refinement paragraph unclear

(*) cannot conclude causation, only association

3/4.25

R1: Group 42 9.25/12 → 4.625/6

1. Formatting:

0.75/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **7** pages
- informative title
- member names on all pgs
- all pages numbered
- no blurry plots (**NOT png**)

too many digits (p-values)

2. Introduction/Background:

1/1

- brief statement of scientific question
- all variables defined (new paragraph)

1.75/2

3. EDA:

- univariate numerical
- univariate graphical
- bivariate numerical (cor)
- bivariate graphical — SQUARE pairs plots

- Don't need figures 3+4

1.25/2

4. Model fitting:

First write out the mathematical model

- state how model fitted (ie, LS)
- CLEARLY describe how model selected
- define all terms

not model 'validity'

1.75/2

5. Model assessment:

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):
- qq normal plot of residuals, — SQUARE
- residuals vs. fitted

- QQ not on diagonal — compared to a reference line (see documentation for qqline())

6.5/7.75

0.5/1

6. Write out final estimated model **mathematically** — no error

hat on response variable — max **2 sig digits** on coefs

0.75/1.25

7. Plots:

label size (not too small) — placement — (captions) – more informative

NOT BLURRY

0.5/1

8. Conclusions – use paragraphing

recap analysis — state main findings (careful with interpretation)

1/1

9. Language quality:

poor — satisfactory — good — excellent

10. Other comments:

2.75/4.25

R1: Group 45 6.5/12 → 3.25/6

1. Formatting:

+Don't need my name

all margins 2.5cm

0.75/0.75 12 pt size

no raw R code or output

max **7** pages

informative title

member names on all pgs

all pages numbered

no blurry plots (**NOT png**)

2. Introduction/Background:

1/1

brief statement of scientific question

all variables defined

3. EDA:

1/2

univariate numerical

univariate graphical

bivariate numerical (cor)

bivariate graphical

4. Model fitting: → why only doing simple reg ??

0.5/2

state how model fitted (ie, LS)

CLEARLY describe how model selected

define all terms

Do mult reg

use numerical methods

5. Model assessment: very incomplete

0.25/2

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):

qq normal plot of residuals, ← SQUARE

residuals vs. fitted

- plot interpretations? Make clear

3.5/7.75

(ok given errors)

6. Write out final estimated model **mathematically** 1/1

hat on response variable max **2 sig digits** on coefs

0.75/1.25 7. Plots: + shapes

label size (not too small) captions

placement **NOT BLURRY**

0.25/1 8. Conclusions

recap analysis state main findings

Vague and generic

9. Language quality: 1/1

poor satisfactory good excellent

10. Other comments:

3/4.25

R1: Group 70 7.5/12 → 3.75/6

1. Formatting: 0.75/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **7** pages
- informative title
- member names on all pgs
- all pages numbered
- no blurry plots (**NOT png**)

2. Introduction/Background: 1/1

- brief statement of scientific question
- all variables defined

3. EDA: 1/2

- univariate numerical
- univariate graphical → histograms (not boxplots)
- bivariate numerical (cor)
- bivariate graphical pairs plots

4. Model fitting: 0.75/2

First write out mathematical model

- state how model fitted (ie, LS)
- CLEARLY describe how model selected
- define all terms

5. Model assessment: 1/2

- CLEARLY state model assumptions:
  1. errors have mean 0
  2. errors are homoscedastic (same variance)
  3. errors are uncorrelated
  4. errors are normally distributed
- carry out assessment (graphics):
  qq normal plot of residuals,
  residuals vs. fitted

'assess' (not 'ensure') normality

- plot interpretations?
- you assess not 'validate'

4.5/7.75

0.5/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

1/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.5/1

8. Conclusions

recap analysis state main findings

1/1

9. Language quality:

poor satisfactory good excellent

10. Other comments:

- log trans 'necessitated'? unclear

3/4.25

R1: Group 71 5.5/12 → 2.75/6

1. Formatting: 0.75/0.75

all margins 2.5cm

12 pt size

no raw R code or output

max **7** pages

informative title

member names on all pgs

all pages numbered

no blurry plots (**NOT png**)

2. Introduction/Background: 1/1

brief statement of scientific question

all variables defined

3. EDA: 0.75/2 – Don't need table 3

univariate numerical

univariate graphical

bivariate numerical (cor)

bivariate graphical

– hists for all vars ← plots too small

4. Model fitting: 0.5/2

~~First write the mathematical model~~

state how model fitted (ie, LS)

CLEARLY describe how model selected

define all terms

very incomplete

5. Model assessment: 0.75/2

↳ check that your comments are correct

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

very vague

carry out assessment (graphics):
qq normal plot of residuals, – square
residuals vs. fitted

– The plots don't check, you use the plots to assess...

– you assess not 'validate' assumptions

3.75/7.75

0.25/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

0.5/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0/1

8. Conclusions

recap analysis state main findings

very generic and vague

1/1

9. Language quality:

poor satisfactory good excellent

10. Other comments:

- exploratory data analysis
- layout not good - too much blank space

1.75/4.25

R1: Group 82 7/12 → 3.5/6

1. Formatting:

0/0.75

all margins 2.5cm

12 pt size throughout

no raw R code or output

max ~~7~~ pages

informative title

member names on all pgs

all pages numbered

no blurry plots (**NOT png**)

1/1 2. Introduction/Background:

brief statement of scientific question

all variables defined

1/2 3. EDA: Don't need Figure 2

univariate numerical

univariate graphical

bivariate numerical (cor) (all pairs) matrix

bivariate graphical

1.5/2 4. Model fitting:

- math model missing error term

state how model fitted (ie, LS)

CLEARLY describe how model selected

define all terms

0.75/2 5. Model assessment: you don't 'verify' assumptions

CLEARLY state model assumptions:

1. errors have mean 0
2. errors are homoscedastic (same variance)
3. errors are uncorrelated
4. errors are normally distributed

carry out assessment (graphics):

qq normal plot of residuals, - SQUARE

residuals vs. fitted

- carefully and clearly interpret plots

4.25/7.75

0.5/1

6. Write out final estimated model **mathematically**

hat on response variable max **2 sig digits** on coefs

1/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.25/1

8. Conclusions

- use paragraphs

recap analysis state main findings

Vague + generic → be specific

1/1

9. Language quality:

poor satisfactory good excellent

10. Other comments:

- why not try logs?

2.75/4.25