Physical Strength & Fear-Related Personality

CMPS-320-01

Andrew Nemkov, Anthony Simone, Claudia Iwinski, Elijah Ihim, Nick Nemkov

Data Source and Background

- The source for our dataset is Kaggle.com, where we were primarily focused on data related to psychology.

- The **Physical Strength & Fear-Related Personality** dataset includes demographic as well as physical strength measures:
  - ➢ Such as **age, gender, ethnicity, grip strength, chest strength**, and most importantly different scores relating to participants' **fearfulness, anxiety, emotional dependence, and sentimentalism.**

- The background of this dataset comes from five university samples all based in the United States. All of the samples are recorded from undergraduate students from those schools, with the last update to the dataset being only 10 months ago.

HEXACO-100 Model

- The dataset includes personality trait items measured using the HEXACO-100 model (widely used and respected in the field of psychology), and each item is scored on a **5-point scale**.
  - ➔ Ranging from **1 (Strongly Disagree)** to **5 (Strongly Agree)**

- Consists of **six major personality dimensions**:
  - ➔ Honesty-Humility (H), **Emotionality (E)**, eXtraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O).

- In the context of the HEXACO-100 items, the focus is on specific traits as seen below:

Fearfulness	*hex_5*: I would feel afraid if I had to travel in bad weather conditions.
Anxiety	*hex_11*: I sometimes can't help worrying about little things.
Emotional Dependence	*hex_17*: When I suffer from a painful experience, I need someone to make me feel comfortable.
Sentimentalism	*hex_23*: I feel like crying when I see other people crying.

HEXACO-100 Items

Fearfulness

hex_29: I don't mind doing jobs that involve dangerous work.

hex_53: When it comes to physical danger, I am very fearful.

hex_77: Even in an emergency I wouldn't feel like panicking.

Emotional Dependence

hex_41: I can handle difficult situations without needing emotional support from anyone else.

hex_65: Whenever I feel worried about something, I want to share my concern with another person.

hex_89: I rarely discuss my problems with other people.

Anxiety

hex_35: I worry a lot less than most people do.

hex_59: I rarely, if ever, have trouble sleeping due to stress or anxiety.

hex_83: I get very anxious when waiting to hear about an important decision.

Sentimentalism

hex_47: When someone I know well is unhappy, I can almost feel that person's pain myself.

hex_71: I feel strong emotions when someone close to me is going away for a long time.

hex_95: I remain unemotional even in situations where most people get very sentimental.

Scientific Questions

1. What correlations exist between psychological scores and strength metrics in the dataset?

2. Can we develop a model that accurately predicts grip strength based on personality traits such as fearfulness, anxiety, emotional dependence, and sentimentalism (as measured by the HEXACO personality model)?

3. Will changing the grip strength to categorical data (via thresholds) significantly improve the predictive power of our model?

4. Do any of the emotional personality traits, as measured by the HEXACO personality model, have any predictive value in regards to grip strength (and therefore to overall physical strength)?

Data Preprocessing - 1

Head of Dataset (First 5 rows)

	age	female	grip	hex_5	hex_11	hex_17	hex_23	hex_29	hex_35	hex_41	hex_47	hex_53	hex_59	hex_65	hex_71	hex_77	hex_83	hex_89
0	23.0	0	34.0	1.0	3	3	1	1	1	3	1	4.0	1.0	3	3.0	3.0	3	4
1	19.0	0	30.0	4.0	4	2	1	3	2	4	5	3.0	4.0	1	5.0	4.0	3	5
2	18.0	1	20.0	4.0	5	4	5	4	2	2	5	1.0	2.0	4	5.0	2.0	4	1
3	19.0	0	38.0	2.0	1	2	3	4	5	4	4	2.0	5.0	2	3.0	4.0	2	4
4	19.0	0	39.0	3.0	2	4	2	2	3	4	3	2.0	1.0	4	4.0	4.0	4	2

	age	female	grip	hex_5	hex_11	hex_17	hex_23	hex_29	hex_35	hex_41	...	hex_65
count	997.000000	997.000000	997.000000	997.000000	997.000000	997.000000	997.000000	997.000000	997.000000	997.000000	...	997.000000
mean	19.292879	0.553661	29.144002	3.063190	3.925777	3.290873	2.996991	3.076229	2.961886	3.195587	...	3.600802
std	1.677419	0.497362	11.914692	1.225162	1.037697	1.150191	1.239814	1.186523	1.305899	1.174897	...	1.046662
min	18.000000	0.000000	5.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000
25%	18.000000	0.000000	20.140000	2.000000	4.000000	2.000000	2.000000	2.000000	2.000000	2.000000	...	3.000000
50%	19.000000	1.000000	27.000000	3.000000	4.000000	4.000000	3.000000	3.000000	3.000000	3.000000	...	4.000000
75%	20.000000	1.000000	37.360000	4.000000	5.000000	4.000000	4.000000	4.000000	4.000000	4.000000	...	4.000000
max	38.000000	1.000000	100.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	...	5.000000

8 rows × 23 columns

Data Description Table
←
Age Range
(18yo-38yo)

Average Grip
(29.14)

Grip Range
(5-100)

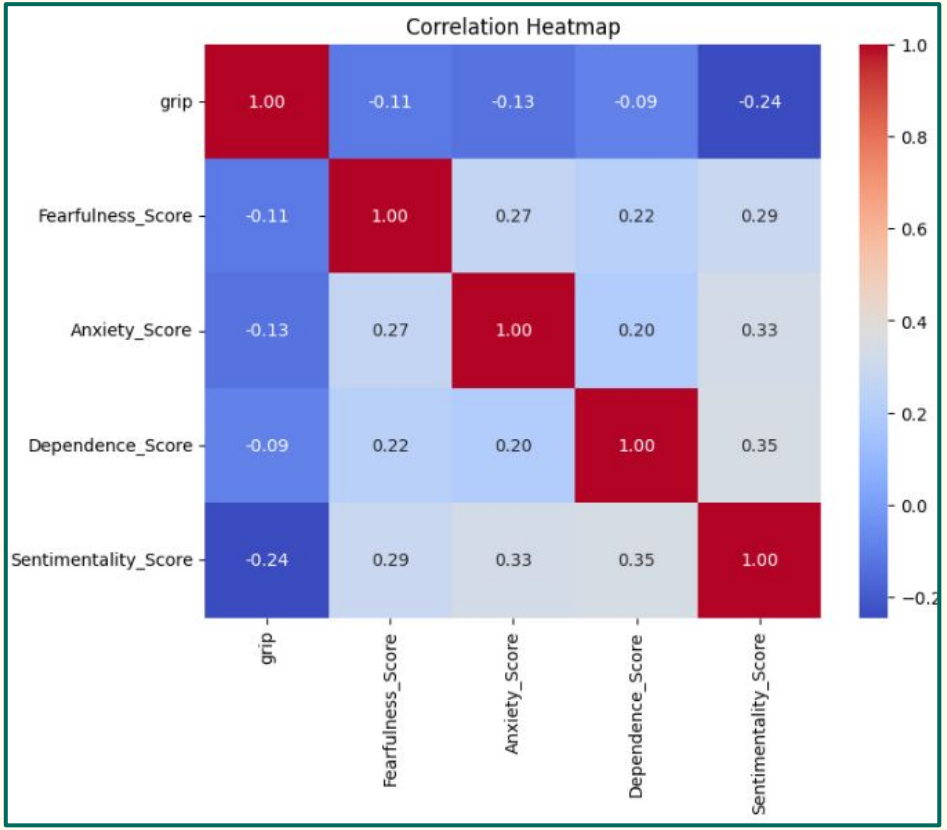
Data Preprocessing - 2

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 997 entries, 0 to 1014
Data columns (total 23 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   age                   997 non-null    float64
 1   female                997 non-null    int64
 2   grip                  997 non-null    float64
 3   hex_5                 997 non-null    float64
 4   hex_11                997 non-null    int64
 5   hex_17                997 non-null    int64
 6   hex_23                997 non-null    int64
 7   hex_29                997 non-null    int64
 8   hex_35                997 non-null    int64
 9   hex_41                997 non-null    int64
 10  hex_47                997 non-null    int64
 11  hex_53                997 non-null    float64
 12  hex_59                997 non-null    float64
 13  hex_65                997 non-null    int64
 14  hex_71                997 non-null    float64
 15  hex_77                997 non-null    float64
 16  hex_83                997 non-null    int64
 17  hex_89                997 non-null    int64
 18  hex_95                997 non-null    int64
 19  Fearfulness_Score     997 non-null    float64
 20  Anxiety_Score         997 non-null    float64
 21  Dependence_Score      997 non-null    int64
 22  Sentimentality_Score  997 non-null    float64
dtypes: float64(10), int64(13)
memory usage: 186.9 KB
```

- The following visualization is a result of implementing the **info** function on our dataset.
- The size of our dataframe is reasonable with 997 observations and 22 features.
- The type of all variables is numerical, with a variety of integer and decimal values.

Correlations within the Dataset

- Relationship between 2 continuous variables
  - (**grip** = 5-100 ; **score** = 1-5)
- Grip strength appears to be negatively impacted by these HEXACO scores
- A high HEXACO score in one area results in a high HEXACO score in another
  - **More anxious == more fearful**; ect.
- **Note:** Gender plays a big role in our dataset. To confirm findings of heat map we need to dive deeper into analysis.
- **Note:** Focus on positive and negative (in psychology high correlation are not likely)



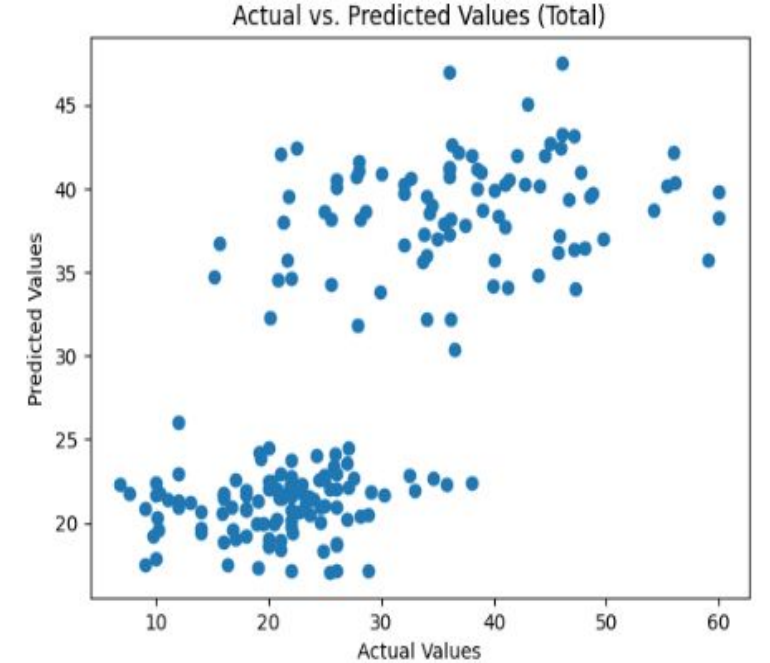
What models did we use?

- **Regression models** (combined, male, and female)
  - **Random Forest**
  - **Lasso**
  - **Ridge**

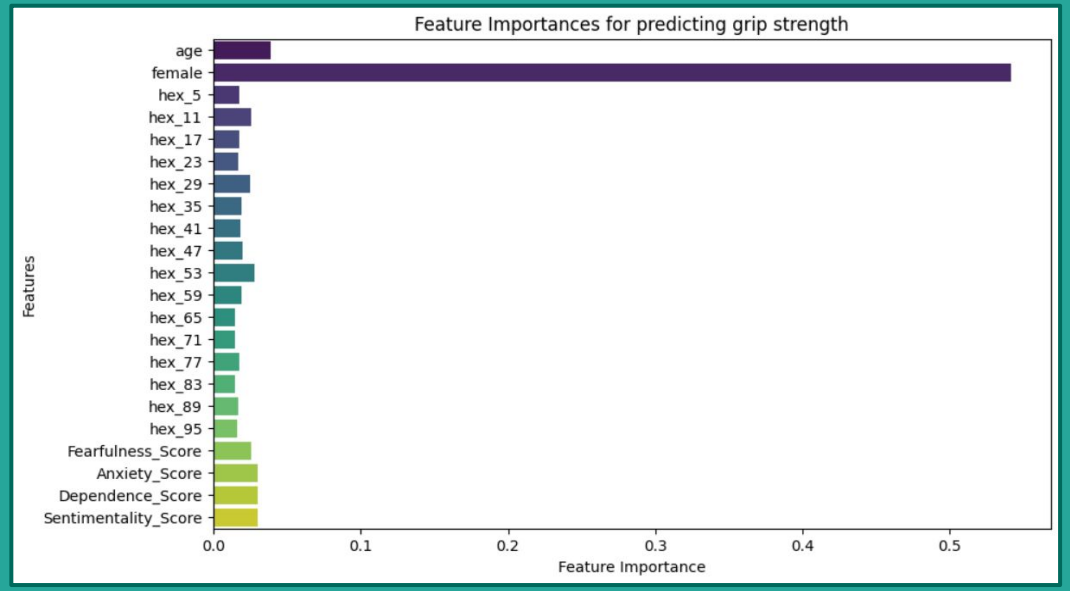
- **Classification models** (male and female)
  - **Random Forest**
  - **Logistic Regression**
  - **Bagging**

- **Clustering models** (combined, male, and female)
  - **K-means Clustering**

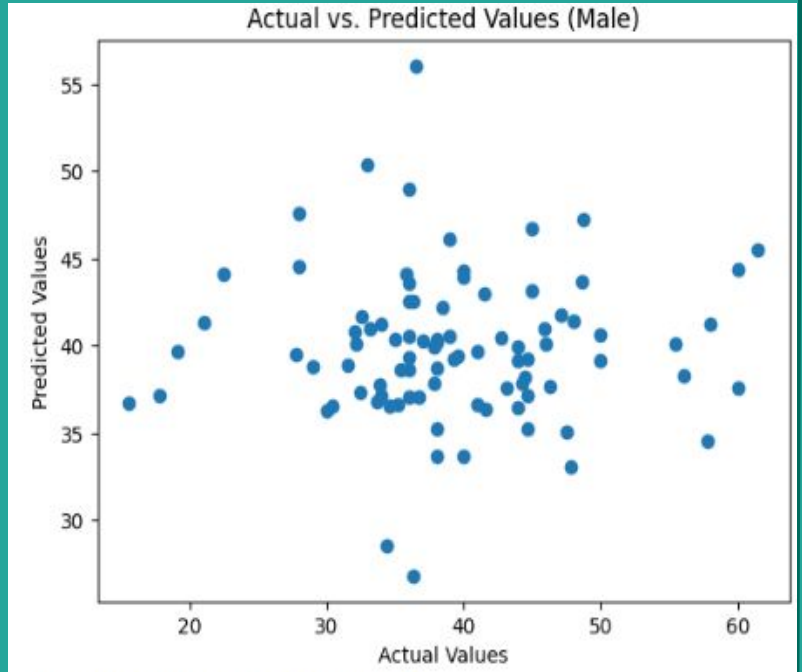
Regression Models - Both Genders



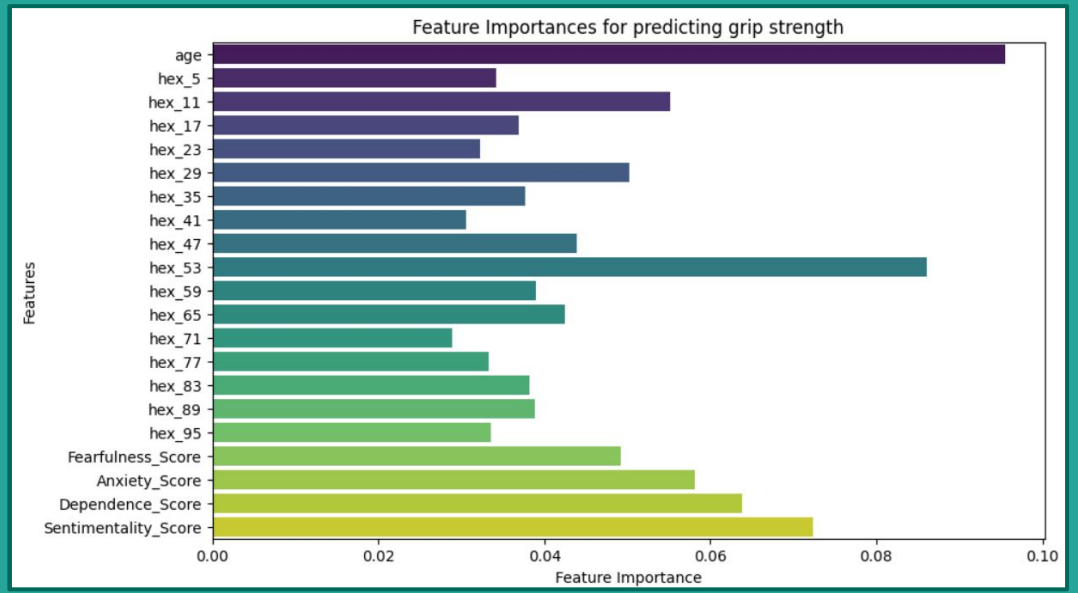
Mean Squared Error: 66.39736513003334
R2 Value: 0.506347182707761



Regression Models - Males



Mean Squared Error: 97.01340692456614
R2 Value: -0.21398166116713124



Models Male DF Regression Results

Model	(Best) Lead Parameter Val	MSE 0 - ∞	R2 -∞ - 1
Random Forest	N_estimators 400	97.01340692456	-0.21398166116
Ridge Regression	Alpha = 2.0	90.58065658197	-0.13348514842
Lasso Regression	**Alpha = 0.5**	**79.55001801169**	**0.004547246889**

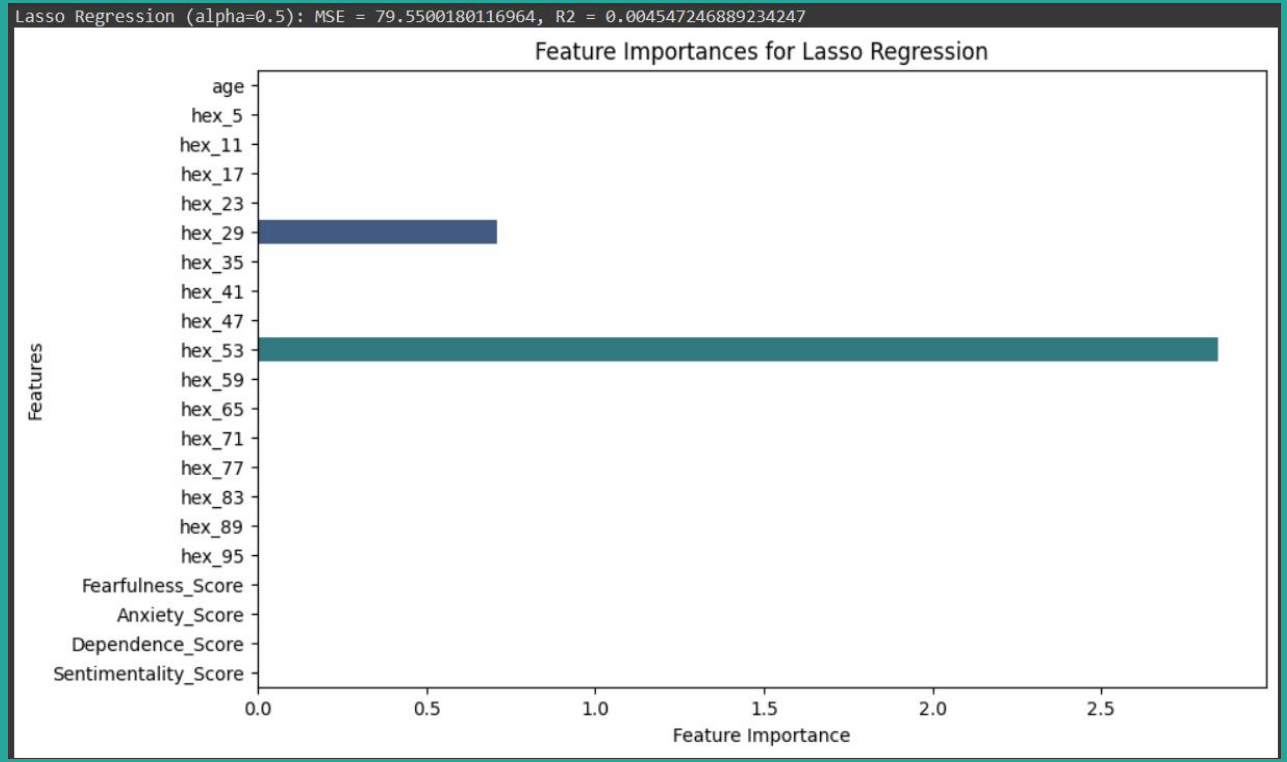
Lasso Feature Importances

Hex_29

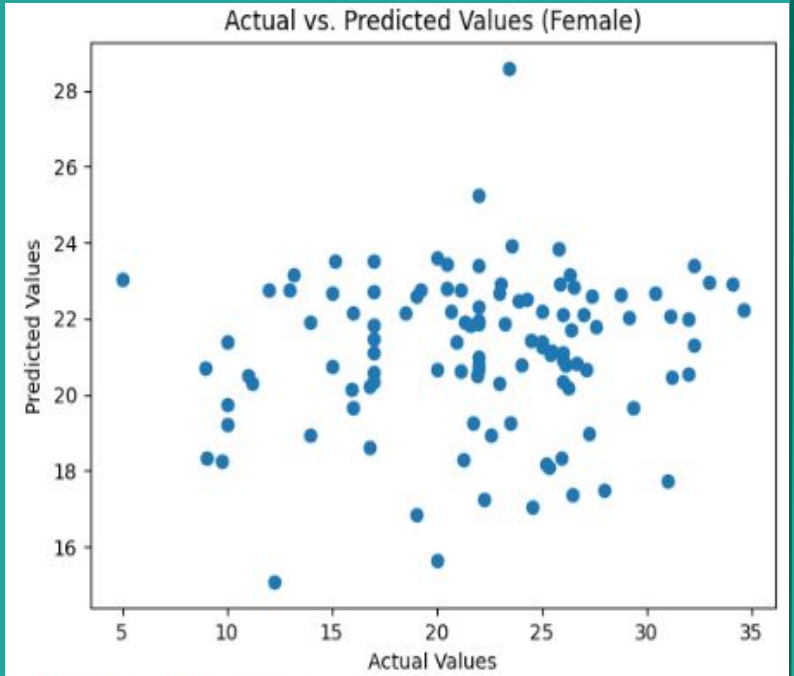
- "I don't mind doing job that involve dangerous work."

Hex_53

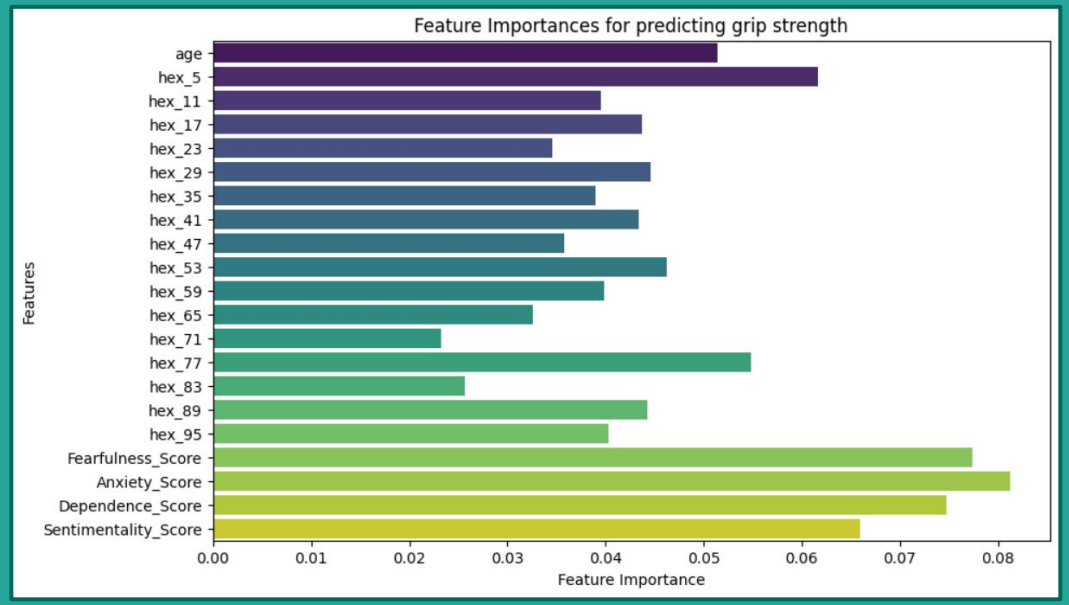
- "When it comes to physical danger, I am very fearful."



Regression Models - Females



Mean Squared Error: 40.611093232004464
R2 Value: -0.04872915167957559



Models Female DF Regression Results

Model	Lead Parameter Val	MSE 0 - ∞	R2 -∞ - 1
Random Forest	N_estimators 100	40.566516586936	-0.04757801726
Ridge Regression	Alpha = 2.0	40.062765862355	-0.03456930392
Lasso Regression	**Alpha = 0.5**	**39.226156756812**	**-0.01296495182**

Classifier Models' Performance - Male Gender

Model	Accuracy	F1 Score
Random Forest	0.4943820224719101	0.440219463752088 76
Logistic Regression	0.5168539325842697	0.460794974024404 95
Bagging	0.438202247191011 25	0.441384632845307 05

- To determine the effectiveness of classifying the dataset, we fit three machine learning models and compared each model's performance.

- Although the accuracy and F1 score for all three models are poor, we can see that logistic classification performed best compared to the other two models.

Classifier Models' Performance - Female Gender

Model	Accuracy	F1 Score
Random Forest	0.43243243243243246	0.3582909832909833
Logistic Regression	0.5045045045045045	0.4286641091773409
Bagging	0.43243243243243246	0.43701375540998183

- Female classification performed similarly to the male classification.
- As seen from the table, accuracy and F1 score for all three models are poor, although logistic classification performed best compared to the other two models.

Classification Performance Reasoning

- *So why does the logistic classification model perform better than random forest and bagging classification for both genders?*

1. Logistic classification is a simpler model compared to random forests and bagging, so it might be better suited for our dataset with only 20 features.

2. Also, the random forest and bagging models excel in capturing complex patterns in data. Since our dataset does not include any significant complex patterns, the logistic model performs better in classifying this data.

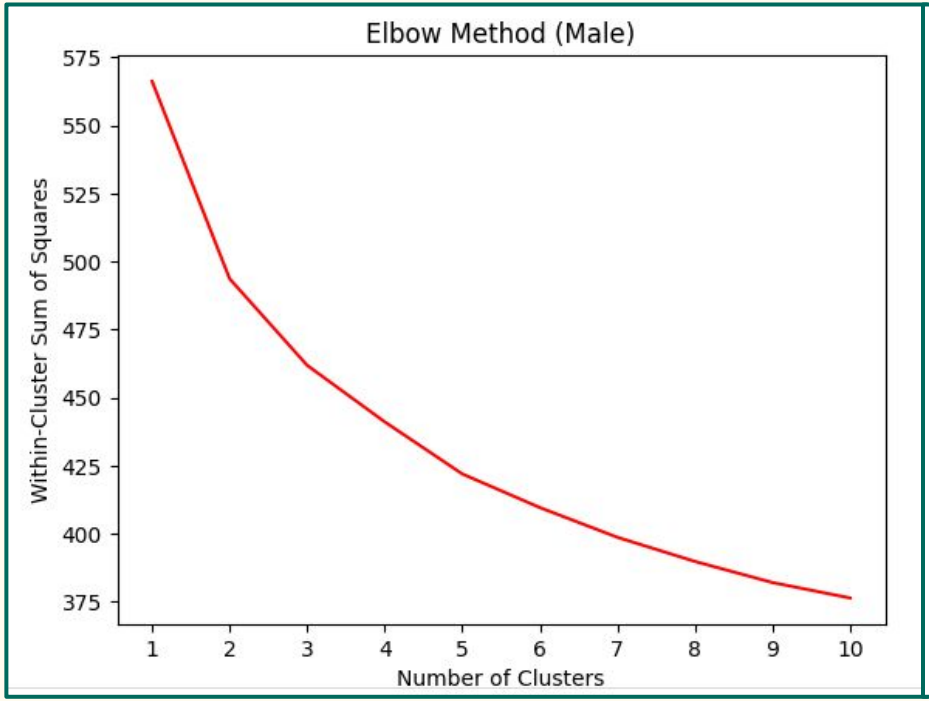
Clustering Data Preparation - 1

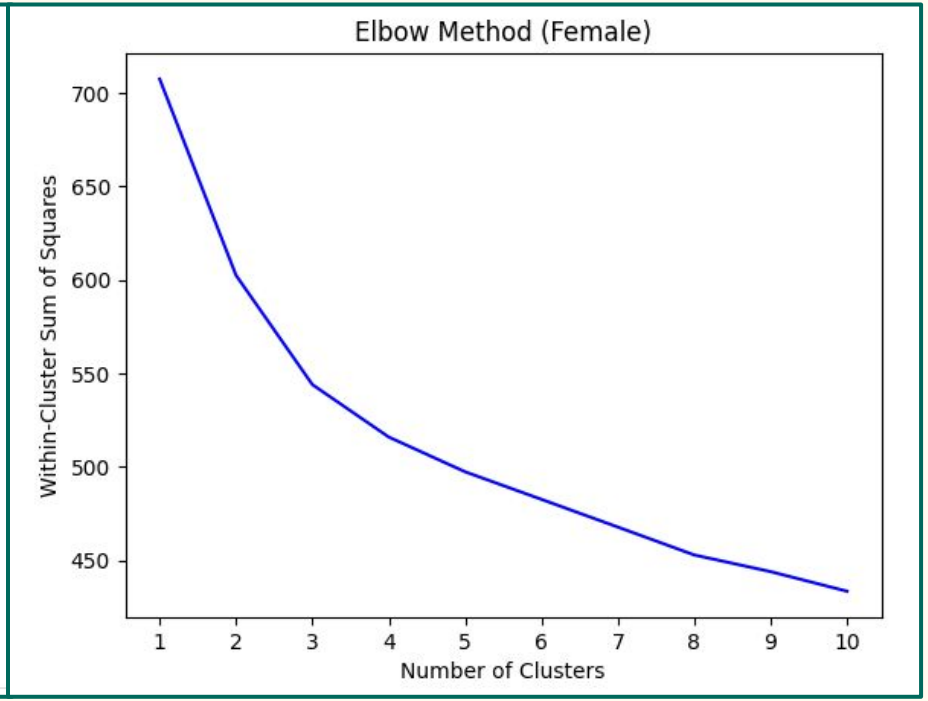
- **Goal:** Split observations into specific number of clusters, based on scores on HEXACO questions.

- **Technique**: K-Means Clustering
  - Must first define value of k via elbow plots

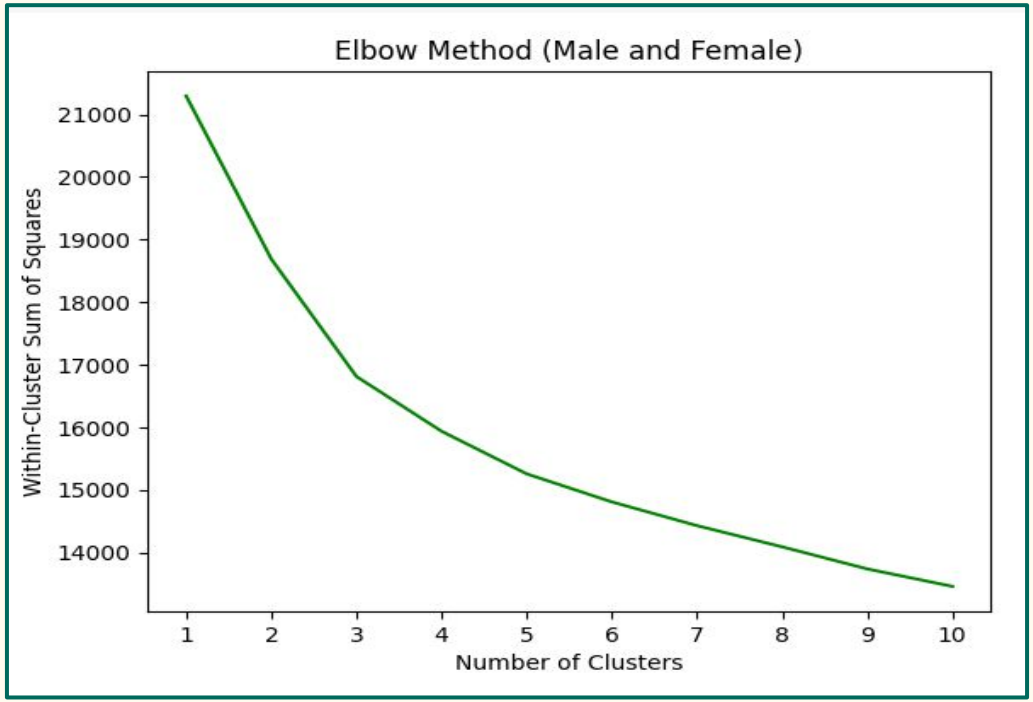
- See if any of the emotional personality traits predict a person's grip strength.

Clustering Data Preparation - 2





Clustering Data Preparation - 3



Clustering Results

Explanation of our process:

- We ran K-means clustering for each individual personality trait (fearfulness, anxiety, dependency, and sentimentalism).

- We plotted the clusters vs their grip strengths.

- We also printed the means of each group, and compared it to the mean of the entire dataset.

- We did that for each of the 4 traits.

- Luckily, with this unsupervised approach, we obtained some meaningful results.

Clustering Results

```
Average of grip strength: 29.1440020060018052
```

Mean Grip Strength for each Cluster (**Fearfulness**):

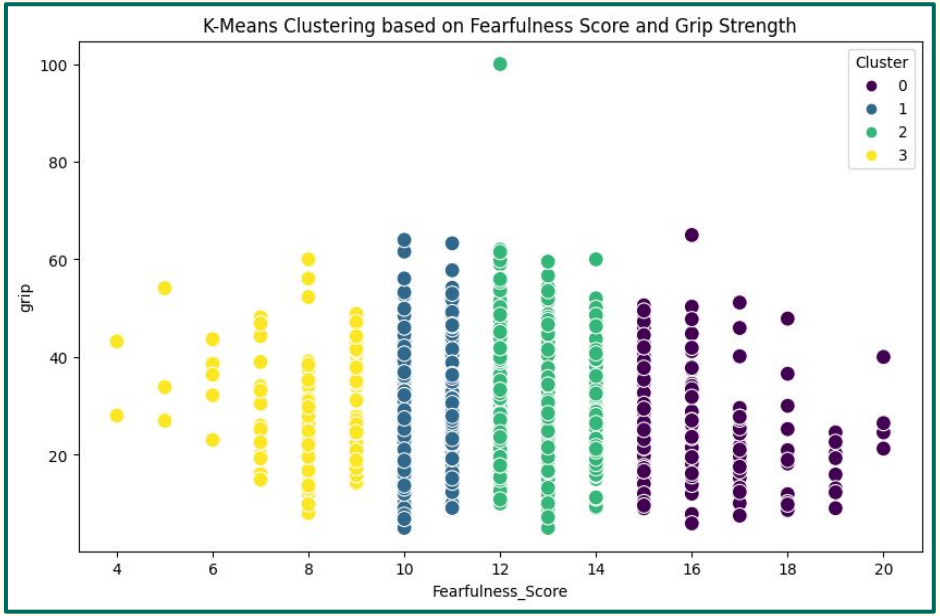
Yellow: 28.403740 **Blue:** 30.172509

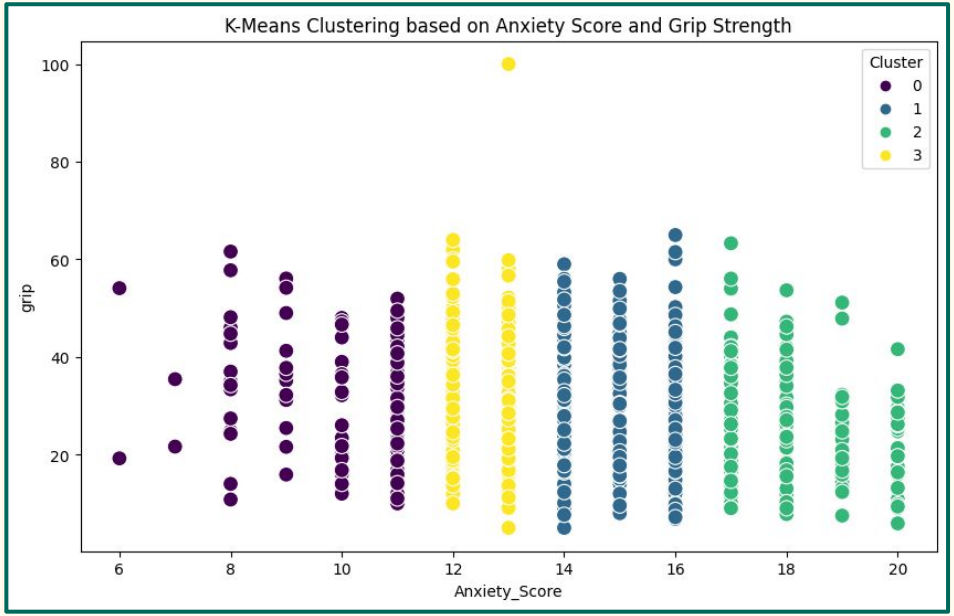
Green: 30.265753 **Purple:** 25.603956

Mean Grip Strength for each Cluster **(Anxiety)**:

Purple: 30.448281 **Yellow:** 29.482445

Blue: 29.848439 **Green:** 25.997558





Clustering Results

```
Average of grip strength: 29.144002006018052
```

Mean Grip Strength for each Cluster **(Dependence)**:

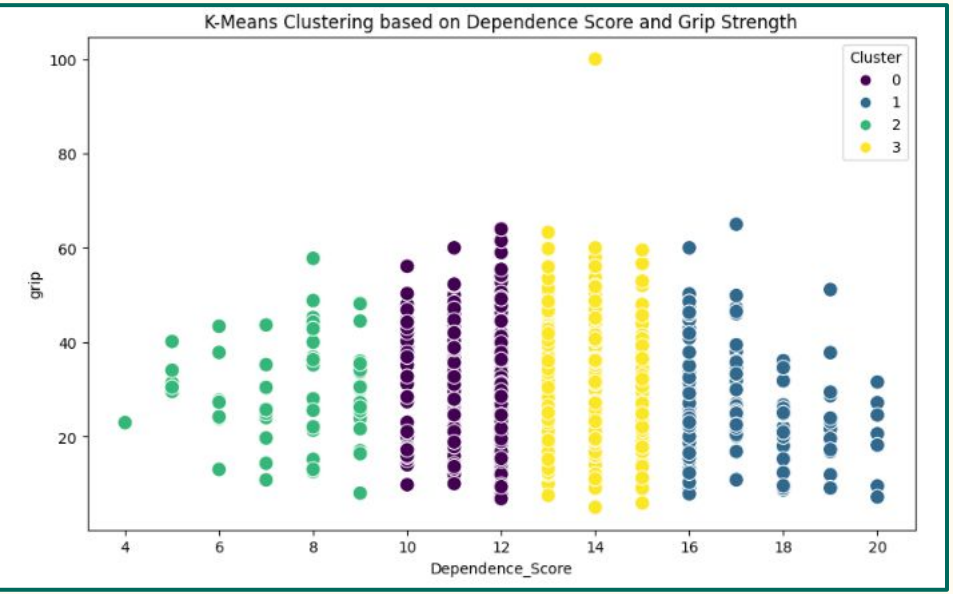
Green: 29.168730 **Purple:** 29.657360

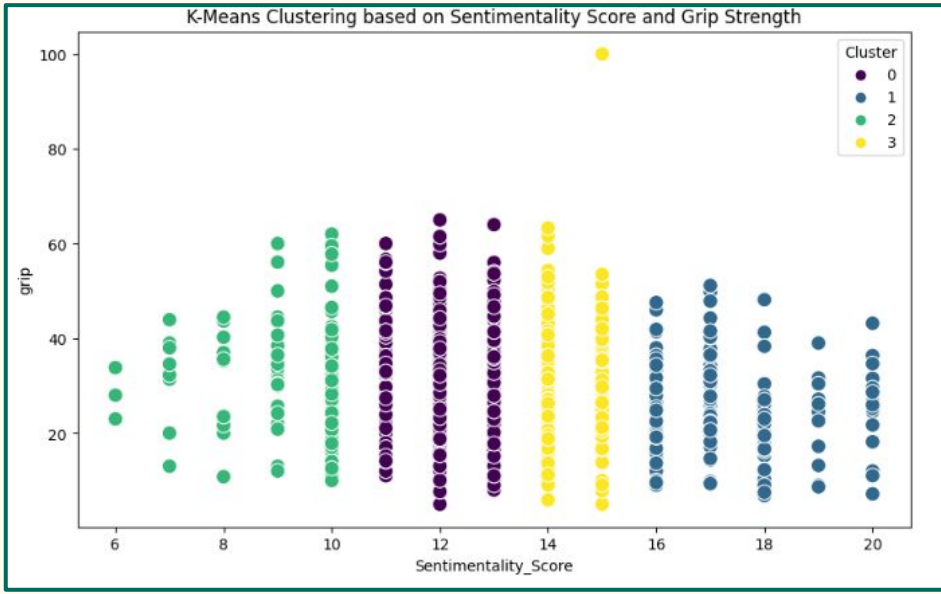
Yellow: 29.689564 **Blue:** 26.416732

Mean Grip Strength for each Cluster **(Sentimentality)**:

Green: 32.460476 **Purple:** 32.008391

Yellow: 27.975768 **Blue:** 24.390221





Combined DF– Main takeaways from results

Trait	**High Strength**	**Low Strength**
Fearfulness	Mid Fearfulness	High Fearfulness
Anxiety	Low anxiety	High anxiety
Dependency	No real impact	High Dependency
Sentimentalism	Low Sentimentalism	High Sentimentalism

Male DF– Main takeaways from results

Trait	**High Strength**	**Low Strength**
Fearfulness	Mid Fearfulness	Extremes(low & high)
Anxiety	Low anxiety	High anxiety
Dependency	Mid Dependency	High Dependency
Sentimentalism	Mid Sentimentalism	High Sentimentalism

Female DF– Main takeaways from results

Trait	**High Strength**	**Low Strength**
Fearfulness	Mid Fearfulness	Extremes(low & high)
Anxiety	No real impact	High Anxiety
Dependency	Low dependency	No real impact
Sentimentalism	No real impact (at all)	No real impact (at all)

Conclusion

- Using regression and classification, we saw very little predictive value in our models. Our models were unable to predict grip strength given personality traits (with any reasonable degree of accuracy)

- Using clustering, however, we were able to extract some meaningful information from our data. We could create clusters based on personality traits, and look at trends between the clusters (their mean grip scores) to draw conclusions.

Conclusion

- **What correlations exist between psychological scores and strength metrics in the dataset?**

  There is a **positive correlation** between **HEXACO scores** to one another, shown when one score is high their counterpart is also high. For **grip strength** in comparison there is an overall **slightly negative** correlation between grip strength and HEXACO scores.

- **Can we develop a model that accurately predicts grip strength based on personality traits such as fearfulness, anxiety, emotional dependence, and sentimentalism (as measured by the HEXACO personality model)?**

  No, our models were generally **inaccurate** when **predicting** grip strength from fear based HEXACO personality traits.

Conclusion

- **Will changing the grip strength to categorical data (via thresholds) significantly improve the predictive power of our model?**

  No. When we changed grip strength to categorical **via thresholds** there was **no noticeable improvement** in the predictive power of our classification models.

- **Do any of the emotional personality traits- as measured by the HEXACO personality model- have any predictive value in regards to grip strength (and therefore to overall physical strength)?**

  Yes. The HEXACO personality traits fearfulness, anxiety, dependency, and sentimentality have **some** predictive value in regards to overall physical strength. A person with **high scores** on the emotional traits tends to have a **lower grip strength** compared to the average person.

Group Member Tasks

- **Elijah Campbell-Ihim** - (Regression models | Clustering vs Grip Strength plots | Finding meaningful trends | Overall project management & organization)
- **Anthony Simone** - (Data Preprocessing | Correlations between psychological scores, and strength metrics | Drawing Conclusions)
- **Nick Nemkov** - (Perform k-means model. | Visualize elbow plots | Visualize feature importances | Regression models)
- **Claudia Iwinski** - (Scoping project and dataset | Initial Random Forest Classifier | Analyzing metrics and results)
- **Andrew Nemkov** - (Logistic and Bagging Classification models | Analyzing metrics and results)

Questions?