**Data Cleaning Agri-PV**

Schritte Datensäuberung:

1. Alle Antworten entfernen, welche nicht im Zeitraum der Umfrage gegeben wurden (Welle 1: ab 03.05.2024 bis 08.06.2024)
2. Alle Antworten entfernen, welche nicht teilnehmen wollten (nur Welle 1, in Welle 2 gab es diese Frage nicht)
3. Alle Antworten in Welle 1 entfernen, welche die Fragen zu den abhängigen Variablen nicht beantwortet haben
4. Alle Antworten in Welle 2 entfernen, welche die Fragen zu Agri-PV nicht vollständig beantwortet haben.
5. Füge vollständige und unvollständige Antworten zusammen
6. Sortiere nach Startdatum
7. Entferne alle Spalten, welche in Welle 2 mit Planted zu tun hatten
8. Füge hinter alle Fragennummern von Welle 2 den Suffix „\_2“ hinzu
9. Füge Welle 1 und Welle 2 zusammen (matching nach „ParticipantID“, ursprünglich „RecipientFirstName“)
10. Alle Fragen zur Ernährung löschen
11. Schnelle Antworten unter 10min entfernen
12. Straightliner entfernen (welche bei Frage Q136 überall das Gleiche angegeben haben) (entfernt wird erst nach Schritt 13)
13. Variablen umbenennen und neu codieren
    1. Geburtsjahre in Alter in Jahren umcodieren
    2. Alter: alle über 90 Jahren entfernen
14. Überprüfung des NIMBY-Treatments
    1. Abgleichen ob die angezeigten Werte für die drei Kreise mit den tatsächlichen Werten übereinstimmt
    2. Dies ist in binärer Variable „circle\_match“ abgebildet
15. Neue Variablen für Regressionen definieren, u. a.
    1. Split variable «after» (1 wenn Antwort nach 24.05.2024 10 Uhr, da wurde der Fehler korrigiert)
16. Subsets für regressionen kreieren. Die Wichtigsten:
    1. «all\_treat» : Alle Antworten (incl. Falsch getreatet) der Treatmentgruppe
    2. «all\_control»: Alle Antworten (incl. Falsch getreatet) der Kontrollgruppe
    3. «all\_match\_correct»: Alle potentiell richtig Getreateten in Kontroll- und Treatmentgruppe
    4. «all\_treat\_correct»: Alle richtig getreateten
    5. “all\_control\_match\_correct”: Kontrollgruppe potentiell richtig getreatet
    6. “all\_match\_correct\_vote”: Alle potentiell richtig Getreateten in Kontroll- und Treatmentgruppe die vor dem 08.06. noch nicht gewählt haben
    7. “data\_after”: Alle Antworten nach 24.05.2024, 10Uhr
    8. «after\_treat»: Treatmentgruppe nach 24.05.
    9. «after\_control»: Kontrollgruppe nach 24.05.
17. Definiere Funktion «model\_control», welche für Regressionen mit Kontrollvariablen genutzt werden kann. Parameter: (dependent\_var, main\_independent\_var, control\_variables, data)
    1. Kontrollvariablen werden im Vektor controls definiert. Aktuell:
       1. age
       2. gender
       3. environment\_score
       4. solar\_open\_space
       5. feelings\_agri\_pv
       6. familiar\_agri\_pv
       7. like\_energy\_agri\_pv