# KPMG AU DATA ANALYTICS SIMULATION

## TASK: 1 - Data Quality Assessment

## Assessment of data quality and completeness in preparation for analysis. Exploring provided datasets.

Dear [Client side Point of Contact],

We appreciate your prompt delivery of the three datasets from Sprocket Central Pty Ltd. The information below presents a summary of the statistics derived from these datasets. Please do not hesitate to inform us if the initial analysis does not match your expectations.

# Exploring Dataset: Customer Demographics

## Initial Analysis:

* Shape: (4000, 13)
* There are no duplicate values

## Blank Fields (Incomplete Data)

Missing values identified in 5 columns including the Default column. They can be treated if required according to the nature of analysis.

* No Last name for following customer Ids:

(4), (67), (106), (139), (197), (211), (250), (251), (257), (275), (356), (460), (475), (494), (514), (526), (657), (660), (676), (684), (690), (703), (744), (780), (790), (857), (860), (916), (927), (938), (1061), (1090), (1129), (1202), (1223), (1253), (1266), (1301), (1320), (1351), (1376), (1453), (1455), (1469), (1512), (1584), (1611), (1651), (1691), (1712), (1727), (1746), (1756), (1858), (1889), (1905), (1935), (1967), (2061), (2073), (2079), (2101), (2137), (2153), (2189), (2311), (2355), (2398), (2407), (2408), (2461), (2473), (2515), (2540), (2580), (2584), (2615), (2630), (2716), (2835), (2894), (2907), (2912), (2918), (2919), (2943), (2945), (2989), (3038), (3052), (3053), (3095), (3137), (3171), (3173), (3180), (3188), (3200), (3259), (3319), (3321), (3324), (3325), (3347), (3364), (3385), (3397), (3401), (3443), (3445), (3447), (3480), (3555), (3597), (3624), (3635), (3651), (3718), (3756), (3817), (3885), (3916), (3927), (3962), (3999)

* No DOB for following customer Ids:

(144), (168), (267), (290), (451), (453), (454), (480), (513), (526), (548), (582), (599), (680), (685), (799), (839), (883), (892), (950), (975), (983), (996), (1038), (1044), (1082), (1174), (1210), (1244), (1351), (1477), (1509), (1583), (1628), (1683), (1740), (1773), (1780), (1806), (1918), (1938), (1990), (2000), (2021), (2069), (2165), (2205), (2252), (2295), (2335), (2341), (2414), (2426), (2469), (2540), (2642), (2647), (2696), (2697), (2854), (2920), (2963), (2999), (3012), (3086), (3151), (3222), (3223), (3255), (3288), (3298), (3312), (3322), (3343), (3365), (3473), (3510), (3513), (3565), (3654), (3718), (3727), (3779), (3883), (3931), (3935), (3998)

* No Job title for following customer Ids:

(4), (6), (7), (11), (22), (23), (30), (46), (52), (60), (62), (74), (81), (108), (110), (161), (167), (178), (185), (197), (207), (217), (229), (244), (262), (267), (276), (288), (296), (302), (319), (336), (338), (340), (344), (368), (370), (379), (384), (388), (391), (398), (405), (412), (418), (422), (434), (442), (443), (462), (464), (466), (479), (482), (497), (505), (507), (510), (519), (522), (535), (539), (541), (548), (550), (557), (567), (577), (614), (616), (622), (627), (638), (659), (671), (680), (683), (691), (693), (700), (703), (711), (712), (714), (731), (733), (738), (740), (741), (745), (758), (765), (766), (771), (773), (786), (787), (812), (815), (820), (827), (828), (860), (864), (871), (874), (879), (882), (887), (895), (901), (903), (908), (913), (944), (953), (981), (983), (992), (1007), (1020), (1023), (1026), (1033), (1034), (1039), (1041), (1045), (1053), (1068), (1071), (1077), (1091), (1112), (1119), (1122), (1123), (1133), (1141), (1145), (1187), (1201), (1218), (1219), (1230), (1233), (1239), (1246), (1249), (1255), (1264), (1268), (1272), (1289), (1292), (1294), (1303), (1315), (1319), (1320), (1335), (1342), (1368), (1374), (1378), (1389), (1393), (1394), (1401), (1417), (1441), (1442), (1443), (1461), (1467), (1471), (1500), (1506), (1507), (1520), (1528), (1548), (1549), (1559), (1560), (1578), (1599), (1604), (1605), (1613), (1614), (1615), (1639), (1643), (1647), (1653), (1678), (1681), (1687), (1694), (1698), (1701), (1706), (1711), (1749), (1751), (1754), (1757), (1769), (1773), (1783), (1792), (1799), (1803), (1811), (1816), (1822), (1849), (1862), (1874), (1880), (1883), (1884), (1892), (1899), (1925), (1948), (1966), (1971), (1972), (1977), (1986), (1994), (2022), (2036), (2052), (2053), (2054), (2071), (2073), (2074), (2083), (2084), (2124), (2146), (2150), (2151), (2159), (2161), (2190), (2208), (2211), (2214), (2230), (2237), (2249), (2250), (2251), (2265), (2266), (2271), (2281), (2300), (2304), (2342), (2407), (2409), (2424), (2426), (2439), (2441), (2446), (2455), (2470), (2483), (2490), (2504), (2506), (2515), (2522), (2532), (2538), (2552), (2553), (2556), (2557), (2558), (2561), (2574), (2581), (2589), (2591), (2613), (2614), (2617), (2632), (2633), (2658), (2660), (2677), (2685), (2703), (2709), (2710), (2721), (2722), (2734), (2737), (2750), (2756), (2757), (2758), (2762), (2764), (2765), (2774), (2776), (2779), (2797), (2801), (2805), (2810), (2815), (2827), (2828), (2833), (2834), (2837), (2842), (2844), (2849), (2866), (2869), (2870), (2901), (2912), (2917), (2918), (2921), (2935), (2941), (2946), (2947), (2949), (2965), (2972), (2986), (2995), (3006), (3008), (3009), (3010), (3021), (3022), (3024), (3039), (3049), (3050), (3051), (3075), (3082), (3083), (3088), (3099), (3106), (3116), (3119), (3122), (3131), (3134), (3147), (3156), (3157), (3159), (3161), (3172), (3182), (3186), (3190), (3207), (3213), (3226), (3227), (3231), (3236), (3250), (3261), (3264), (3267), (3274), (3290), (3304), (3313), (3314), (3318), (3319), (3329), (3330), (3337), (3338), (3342), (3345), (3354), (3372), (3376), (3377), (3378), (3384), (3390), (3391), (3395), (3398), (3410), (3416), (3429), (3436), (3447), (3456), (3457), (3459), (3476), (3479), (3480), (3498), (3499), (3500), (3505), (3516), (3523), (3533), (3539), (3540), (3541), (3548), (3557), (3558), (3563), (3570), (3581), (3584), (3585), (3587), (3590), (3612), (3637), (3642), (3655), (3656), (3657), (3658), (3670), (3672), (3673), (3686), (3689), (3693), (3694), (3705), (3713), (3716), (3720), (3722), (3724), (3733), (3744), (3749), (3754), (3756), (3760), (3762), (3775), (3779), (3784), (3794), (3798), (3804), (3806), (3811), (3822), (3824), (3826), (3827), (3840), (3844), (3865), (3881), (3893), (3909), (3916), (3928), (3929), (3930), (3931), (3933), (3935), (3939), (3945), (3946), (3952), (3959), (3960), (3968), (3972), (3979), (3987), (3999)

* Job industry not mentioned (n/a) for following customer Ids:

(5), (8), (16), (17), (18), (23), (33), (36), (46), (48), (49), (57), (59), (68), (69), (73), (74), (84), (85), (108), (109), (111), (113), (118), (119), (148), (154), (158), (161), (178), (187), (191), (197), (206), (207), (209), (223), (238), (261), (270), (273), (277), (278), (280), (292), (294), (295), (321), (335), (344), (353), (357), (358), (381), (383), (403), (407), (410), (413), (416), (418), (423), (439), (464), (476), (478), (501), (505), (510), (519), (522), (534), (535), (556), (558), (561), (562), (573), (578), (593), (594), (602), (621), (627), (633), (635), (636), (645), (665), (669), (674), (676), (690), (696), (702), (707), (711), (713), (715), (721), (730), (731), (736), (737), (741), (746), (747), (760), (762), (771), (790), (792), (801), (804), (807), (811), (825), (833), (837), (851), (854), (857), (860), (866), (867), (871), (879), (880), (881), (895), (896), (904), (909), (916), (924), (928), (932), (935), (945), (948), (953), (959), (962), (963), (974), (977), (978), (999), (1002), (1008), (1012), (1028), (1029), (1032), (1033), (1034), (1039), (1046), (1048), (1062), (1070), (1073), (1074), (1083), (1088), (1093), (1101), (1105), (1111), (1112), (1115), (1117), (1122), (1126), (1130), (1131), (1137), (1140), (1149), (1158), (1162), (1171), (1172), (1175), (1176), (1183), (1185), (1188), (1191), (1213), (1218), (1220), (1234), (1238), (1239), (1246), (1248), (1262), (1277), (1289), (1292), (1299), (1304), (1308), (1309), (1323), (1329), (1330), (1332), (1336), (1338), (1340), (1348), (1356), (1395), (1409), (1412), (1416), (1436), (1449), (1457), (1464), (1473), (1476), (1480), (1484), (1487), (1488), (1492), (1494), (1498), (1500), (1503), (1512), (1513), (1514), (1539), (1543), (1547), (1562), (1567), (1569), (1574), (1592), (1598), (1599), (1605), (1606), (1608), (1610), (1614), (1618), (1619), (1620), (1622), (1640), (1645), (1650), (1666), (1688), (1691), (1695), (1723), (1742), (1752), (1756), (1764), (1765), (1766), (1769), (1770), (1778), (1781), (1797), (1799), (1803), (1807), (1817), (1823), (1829), (1843), (1844), (1851), (1853), (1854), (1865), (1867), (1870), (1880), (1881), (1893), (1896), (1897), (1903), (1921), (1922), (1924), (1925), (1934), (1935), (1939), (1945), (1946), (1955), (1958), (1960), (1961), (1974), (1981), (1985), (1993), (1995), (1997), (2023), (2025), (2027), (2031), (2036), (2043), (2046), (2050), (2059), (2068), (2070), (2071), (2079), (2080), (2083), (2094), (2101), (2102), (2118), (2131), (2135), (2142), (2143), (2146), (2147), (2149), (2155), (2159), (2163), (2180), (2188), (2191), (2192), (2193), (2200), (2208), (2213), (2219), (2230), (2231), (2232), (2233), (2237), (2242), (2243), (2246), (2259), (2267), (2269), (2273), (2276), (2280), (2281), (2283), (2304), (2310), (2327), (2342), (2346), (2347), (2350), (2352), (2361), (2364), (2376), (2387), (2394), (2395), (2401), (2405), (2406), (2417), (2424), (2425), (2427), (2428), (2430), (2433), (2434), (2447), (2449), (2456), (2461), (2465), (2475), (2485), (2508), (2512), (2518), (2522), (2527), (2537), (2538), (2543), (2548), (2551), (2553), (2554), (2557), (2560), (2563), (2573), (2577), (2585), (2587), (2588), (2589), (2595), (2596), (2616), (2618), (2640), (2641), (2643), (2648), (2652), (2656), (2657), (2664), (2671), (2673), (2683), (2685), (2689), (2700), (2708), (2718), (2729), (2736), (2747), (2776), (2779), (2787), (2799), (2804), (2812), (2818), (2822), (2826), (2841), (2845), (2849), (2850), (2858), (2860), (2867), (2870), (2884), (2886), (2889), (2890), (2893), (2898), (2905), (2912), (2923), (2928), (2939), (2940), (2942), (2943), (2945), (2965), (2969), (2972), (2977), (2987), (2988), (3006), (3010), (3014), (3018), (3021), (3022), (3038), (3043), (3050), (3052), (3055), (3057), (3059), (3081), (3083), (3085), (3092), (3094), (3100), (3107), (3119), (3137), (3148), (3149), (3176), (3177), (3186), (3190), (3199), (3203), (3205), (3207), (3208), (3215), (3231), (3240), (3248), (3252), (3257), (3259), (3266), (3269), (3272), (3276), (3277), (3281), (3287), (3293), (3301), (3304), (3309), (3316), (3321), (3329), (3340), (3344), (3351), (3353), (3359), (3369), (3370), (3381), (3382), (3384), (3389), (3390), (3398), (3411), (3412), (3414), (3419), (3435), (3439), (3443), (3451), (3454), (3455), (3456), (3468), (3480), (3514), (3517), (3521), (3523), (3525), (3530), (3540), (3544), (3545), (3548), (3558), (3562), (3564), (3569), (3576), (3581), (3587), (3592), (3599), (3600), (3605), (3606), (3611), (3619), (3632), (3634), (3640), (3642), (3643), (3647), (3653), (3655), (3657), (3663), (3671), (3679), (3683), (3692), (3697), (3701), (3710), (3720), (3725), (3731), (3733), (3741), (3746), (3748), (3751), (3756), (3757), (3763), (3767), (3773), (3786), (3793), (3800), (3819), (3829), (3835), (3837), (3840), (3842), (3848), (3850), (3852), (3853), (3855), (3860), (3864), (3871), (3877), (3878), (3879), (3887), (3892), (3893), (3896), (3903), (3907), (3911), (3918), (3925), (3929), (3933), (3947), (3951), (3957), (3959), (3963), (3966), (3974), (3976), (3983), (4000)

* No tenure mentioned for following customer Ids:

(144), (168), (267), (290), (451), (453), (454), (480), (513), (526), (548), (582), (599), (680), (685), (799), (839), (883), (892), (950), (975), (983), (996), (1038), (1044), (1082), (1174), (1210), (1244), (1351), (1477), (1509), (1583), (1628), (1683), (1740), (1773), (1780), (1806), (1918), (1938), (1990), (2000), (2021), (2069), (2165), (2205), (2252), (2295), (2335), (2341), (2414), (2426), (2469), (2540), (2642), (2647), (2696), (2697), (2854), (2920), (2963), (2999), (3012), (3086), (3151), (3222), (3223), (3255), (3288), (3298), (3312), (3322), (3343), (3365), (3473), (3510), (3513), (3565), (3654), (3718), (3727), (3779), (3883), (3931), (3935), (3998)

* Common IDs between Data Set 1 and Data Set 2:

Common IDs: (526), (1351), (2540)

* Common IDs between Data Set 1 and Data Set 3:

Common IDs: (4), (197)

## Further identified discrepancies:

* Typing error; Customer ID 54: Femal= Female

1: F= Female

57: M=Male

* Deceased data can be ignored: Customer Ids (3790), (753)
* Age discrepancy for customer ID 34, age 180 years not possible
* Default column values inconsistent, therefore dropping the column

# Exploring Dataset: Transactions

## Initial Analysis:

* Shape: (20000, 13)

There are missing values in 7 columns. These can be treated if required depending upon the nature of analysis.

* Missing online\_order true or false values for following Transaction IDs:

(16725), (19844), (16846), (8866), (15165), (14893), (18847), (6378), (1193), (12109), (4047), (15711), (17435), (577), (2633), (3135), (17415), (16853), (14545), (18472), (6148), (12692), (16161), (10455), (6038), (13631), (10373), (14642), (13057), (8029), (17164), (9057), (5802), (5964), (10857), (16285), (98), (5029), (15190), (477), (18023), (17233), (3454), (10095), (18755), (170), (6471), (14060), (16698), (16102), (12678), (5835), (13408), (13025), (7727), (10478), (2096), (12644), (12753), (9362), (1818), (4889), (11161), (5476), (14316), (1938), (18712), (15242), (19515), (9386), (14575), (7325), (5888), (1622), (18983), (13713), (2390), (11620), (12711), (18069), (15061), (15066), (1981), (15112), (15942), (1553), (4162), (9447), (5756), (4867), (6415), (14296), (2214), (6687), (1922), (7974), (15815), (8453), (4268), (8165), (10176), (11965), (12758), (13439), (7080), (12934), (12320), (6785), (5062), (19370), (11271), (3033), (3533), (11711), (10254), (18625), (2773), (4385), (5434), (3201), (10035), (3958), (530), (19083), (6628), (13284), (12073), (11985), (18011), (7460), (14167), (4754), (8810), (7998), (10501), (14187), (8074), (399), (11156), (12331), (11755), (1061), (17716), (13292), (15825), (19283), (5354), (19636), (1382), (1362), (7257), (2284), (8406), (7777), (16021), (9802), (10368), (2842), (5455), (9996), (9180), (5562), (13837), (6356), (12467), (343), (15382), (4822), (1072), (16347), (12897), (12302), (4051), (14621), (14333), (19434), (6467), (18723), (337), (9690), (16506), (17072), (17335), (5673), (8028), (16255), (4803), (9368), (15115), (5416), (1160), (2648), (5461), (6520), (19574), (2647), (19304), (5978), (9771), (8934), (16187), (7379), (15739), (1355), (11108), (13223), (19581), (19025), (15106), (8037), (251), (6744), (15274), (12403), (5893), (6573), (15963), (10698), (10579), (5743), (10788), (8087), (18607), (15374), (1524), (11698), (7834), (4478), (2618), (18502), (15546), (18752), (12790), (16046), (7529), (918), (16076), (5643), (14795), (12918), (9934), (16729), (17734), (4044), (4832), (7268), (13512), (9658), (13314), (8626), (10074), (770), (761), (3554), (5022), (11016), (15446), (4348), (13706), (16548), (17560), (17061), (12592), (16069), (301), (4344), (3955), (3026), (15030), (1988), (13531), (13212), (6896), (13671), (4071), (19101), (6853), (13357), (14668), (11675), (915), (15758), (1556), (12503), (17085), (14429), (18086), (10299), (12930), (14566), (11324), (13826), (6112), (7676), (16222), (13593), (18336), (3813), (9969), (16022), (10800), (18690), (16919), (7430), (14480), (19272), (15902), (12617), (17106), (19496), (4681), (1888), (5575), (9706), (15366), (4654), (6400), (559), (18043), (12770), (6250), (9819), (14799), (8694), (10437), (9540), (15304), (9345), (167), (16520), (19369), (1201), (10139), (1372), (5718), (12004), (8950), (3905), (14869), (17890), (14762), (10185), (7323), (11913), (8161), (6475), (8116), (8053), (8041), (9301), (12303), (844), (3930), (14761), (17862), (18822), (10029), (2615), (16744), (7955)

* Product ID=0 and data missing for columns: brand, product\_line, product\_class, product\_size, standard\_cost, product\_first\_sold\_date with reference to the following Transaction IDs:

(137), (160), (367), (407), (677), (781), (1004), (1131), (1197), (1283), (1654), (1681), (1753), (1779), (2187), (2303), (2948), (2991), (3195), (3260), (3409), (13025), (3782), (3878), (4025), (4074), (4128), (4223), (4225), (4228), (4339), (4405), (4605), (4653), (4714), (4833), (4896), (4967), (5065), (5159), (5207), (5261), (5267), (5603), (5862), (6072), (6138), (6168), (6251), (6269), (6419), (6430), (6500), (6596), (6644), (6675), (6933), (6961), (6988), (7016), (7047), (7206), (7235), (7356), (7610), (7695), (7730), (7735), (7780), (7851), (7949), (8160), (8216), (8298), (8322), (8329), (8424), (8486), (8527), (8586), (8610), (8611), (8925), (8956), (9187), (9231), (9276), (9326), (9420), (9454), (9479), (9630), (9715), (9760), (10002), (10045), (10541), (10896), (10960), (11116), (11219), (11393), (11518), (11531), (11549), (11834), (11843), (12022), (12072), (12139), (12206), (12323), (12344), (12347), (12462), (12817), (12870), (12912), (12969), (13011), (13028), (13141), (13344), (13345), (13369), (13380), (13460), (13520), (13594), (13660), (13681), (13720), (13739), (13809), (13849), (13873), (14228), (14305), (14410), (14423), (14444), (11016), (14608), (14641), (14755), (14836), (14846), (14922), (15148), (15268), (15529), (15622), (15709), (15928), (16003), (16010), (16033), (16095), (16249), (16289), (16497), (16661), (16682), (16767), (16907), (16935), (17150), (17236), (17292), (17332), (17432), (17535), (17768), (17919), (17988), (18022), (18052), (18245), (18278), (18280), (18345), (18493), (18544), (18554), (18779), (18820), (19018), (19046), (19133), (19177), (19197), (19206), (19341), (19384), (19794), (19860), (19872)

* standard\_cost not in accurate format for following transaction IDs: (8708), (16701), (17469)
* There are no duplicate values
* The values in product first sold date aren’t correct as they imply everything happening the same day at different times

# Exploring Dataset: New Customer List

## Initial Analysis:

* Shape: (1000, 18)
* There are missing values in 4 columns. They can be dropped or treated according to the nature of analysis

These include: (column number of null values)

last\_name 29

DOB 17

job\_title 106

job\_industry\_category 165

* There are no duplicate values
* Gender Unknown for 17 rows

# Exploring Dataset: Customer Address

## Initial Analysis:

* Shape: (3999, 6)
* There are no missing/null values
* There are no duplicate values
* All the columns appear to have consistent information

Going forward, our team will proceed with the data cleaning, standardization, and transformation processes as part of our model analysis efforts. We anticipate that questions may arise during this process, and we will diligently document any assumptions made. Once this phase is completed, we would greatly appreciate the opportunity to collaborate to ensure that all assumptions are in harmony with your company's perspective.

Regards,

Suryadeep Khungar