**基于Kmeans聚类的用户画像分析**

基于调查问卷中的性别、线下社交活跃度等变量，本文应用Kmeans聚类模型把受访者划分为三类，并通过各类群体在多个维度上的平均值来分析其特征，建立用户画像，便于研究者针对性分析和企业针对性营销。

**（一）研究方法**

1.算法原理

Kmeans聚类用于将数据点划分为 K 个簇，使得同一簇内的数据点相似度较高，而不同簇之间的数据点差异较大。该算法的核心思想是通过迭代优化簇中心的位置，以最小化簇内数据点到质心的距离。其结果依赖于K值（分为几类）和分类起始点。

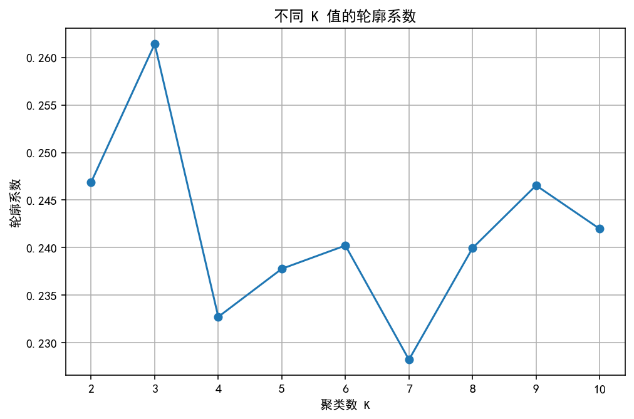
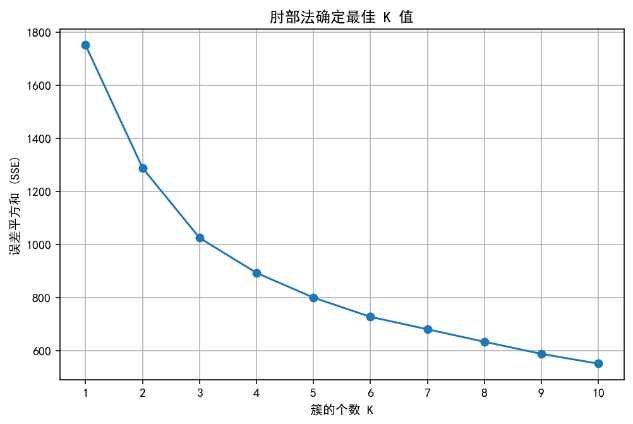
2.变量选取

选取了性别、年龄、身份、月消费额、孤独感、抑郁度六个维度的个人信息作为分类变量，以研究在情感陪伴AI各个维度上的使用行为。

3.选择最佳聚类数

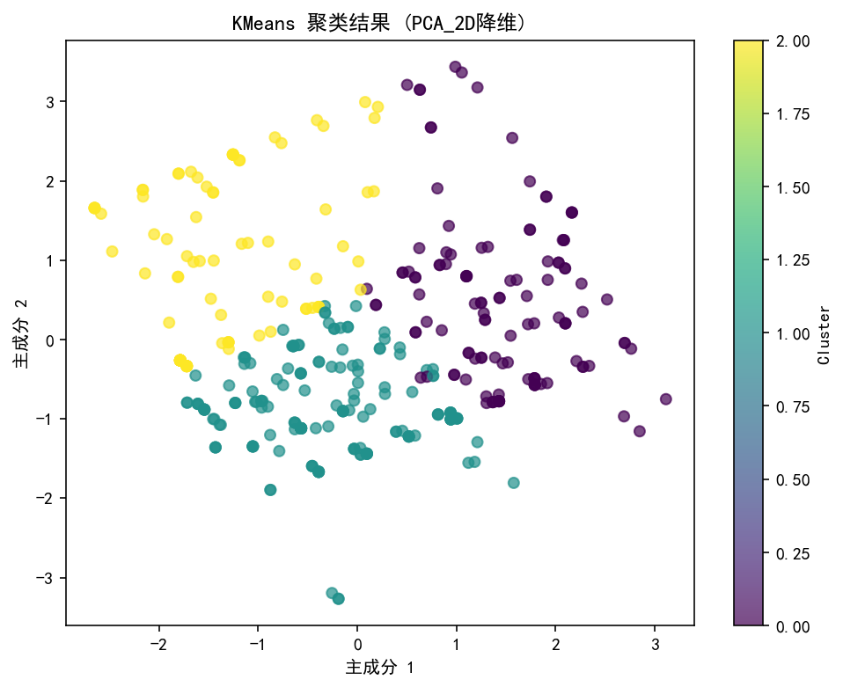
使用肘部法和轮廓系数来选择最佳的聚类数。在研究中，我们尝试了K=2-10的聚类数，结果显示K=3时的聚类效果最优。

图 肘部法、轮廓系数确定K值结果图



3.聚类计算及结果可视化

使用Kmeans算法拟合数据，求得受访者所属的聚类，使用PCA主成分分析方法对数据降维，生成了二维可视化效果，可见聚类本身分布相对集中，各聚类之间分布相对分散，聚类效果较好。

图 Kmeans聚类结果（PCA降维）

**（二）聚类结果分析**

1.聚类结果汇总

表 Kmenas聚类结果

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | 1 | 2 | 3 |
| 聚类中心 | | | |
| **Age** | -0.20 | -0.39 | 0.96 |
| **Spending** | -0.07 | -0.47 | 0.93 |
| **Loneliness** | 0.99 | -0.35 | -0.65 |
| **Depre** | 1.15 | -0.57 | -0.47 |
| **Gender\_Female** | 0.40 | 0.36 | 0.68 |
| **Identity\_graduate** | 0.11 | 0.13 | 0.03 |
| **Identity\_part** | 0.10 | 0.13 | 0.02 |
| **Identity\_work** | 0.26 | 0.16 | 0.89 |
| **Sample\_Count** | 119 | 165 | 92 |
| 聚类均值 | | | |
| **ReasonLonely** | 0.57 | 0.38 | 0.47 |
| **ReasonEmo** | 0.36 | 0.44 | 0.40 |
| **ReasonSupport** | 0.47 | 0.27 | 0.19 |
| **ReasonSocial** | 0.13 | 0.14 | 0.15 |
| **ReasonEnter** | 0.26 | 0.41 | 0.28 |
| **ReasonNovel** | 0.23 | 0.20 | 0.13 |
| **ReasonOther** | 0.00 | 0.02 | 0.00 |
| **ExperOverall** | 3.70 | 3.70 | 3.47 |
| **ExperNature** | 3.51 | 3.84 | 3.62 |
| **ExperSupport** | 3.70 | 3.64 | 3.60 |
| **ExperPrivacy** | 3.60 | 3.69 | 3.38 |
| **ExperPersonal** | 3.62 | 3.66 | 3.40 |
| **ImporNature** | 4.00 | 3.77 | 3.60 |
| **ImporSupport** | 4.09 | 4.02 | 3.79 |
| **ImporPrivacy** | 3.96 | 4.02 | 3.74 |
| **imporPersonal** | 3.91 | 4.06 | 3.57 |
| **ScenarioSleep** | 0.45 | 0.64 | 0.72 |
| **ScenarioWork** | 0.62 | 0.50 | 0.62 |
| **ScenarioDown** | 0.68 | 0.55 | 0.68 |
| **ScenarioNovel** | 0.58 | 0.67 | 0.68 |
| **ScenarioTravel** | 0.43 | 0.42 | 0.68 |
| **ScenarioOther** | 0.00 | 0.00 | 0.02 |
| **NotReasonProduct** | 0.39 | 0.44 | 0.56 |
| **NotReasonNece** | 0.73 | 0.76 | 0.53 |
| **NotReasonPrivacy** | 0.36 | 0.24 | 0.51 |
| **NotReasonOther** | 0.03 | 0.06 | 0.00 |
| **Try** | 2.83 | 2.55 | 2.89 |
| **ConcernSocial** | 0.37 | 0.48 | 0.60 |
| **ConcernMoral** | 0.39 | 0.45 | 0.39 |
| **ConcernPrivacy** | 0.69 | 0.70 | 0.71 |
| **ConcernExper** | 0.40 | 0.34 | 0.26 |
| **ConcernNot** | 0.24 | 0.19 | 0.25 |
| **ConcernOther** | 0.03 | 0.02 | 0.04 |
| **Pay** | 2.81 | 2.38 | 2.80 |
| **Price** | 2.44 | 2.06 | 2.29 |
| **ChooseDialogue** | 0.56 | 0.56 | 0.48 |
| **ChoosePrice** | 0.13 | 0.16 | 0.13 |
| **ChoosePrivacy** | 0.41 | 0.45 | 0.45 |
| **ChoosePersonal** | 0.54 | 0.44 | 0.32 |
| **ChooseBrand** | 0.03 | 0.05 | 0.16 |
| **ChooseUser** | 0.13 | 0.25 | 0.20 |
| **ChooseFunction** | 0.43 | 0.41 | 0.41 |
| **ChooseOther** | 0.02 | 0.01 | 0.01 |
| **Market** | 3.45 | 3.32 | 3.52 |
| **Support\_network** | 0.19 | 0.13 | 0.15 |
| **Support\_self** | 0.38 | 0.33 | 0.55 |
| **Support\_friend** | 0.30 | 0.44 | 0.25 |
| **UsedFreq\_sometimes** | 0.23 | 0.23 | 0.35 |
| **UsedFreq\_heard** | 0.34 | 0.39 | 0.33 |
| **UsedFreq\_never** | 0.21 | 0.22 | 0.16 |
| **PayModel\_use** | 0.16 | 0.11 | 0.14 |
| **PayModel\_subs** | 0.24 | 0.25 | 0.22 |
| **PayModel\_purchase** | 0.31 | 0.25 | 0.30 |
| **PayModel\_other** | 0.01 | 0.02 | 0.00 |
| **Cluster** | 0 | 1 | 2 |
| **Sample\_Count** | 119 | 165 | 92 |

图 聚类中心热力图

图表, 条形图, 树状图

AI 生成的内容可能不正确。

2.群体特征解读及用户画像构建

**第一类群体(119人，占比32%):**

人群特征上，体现为年龄偏小、消费水平中等、孤独感和抑郁程度较高，大多数为在校本科学生，也有一部分参与工作。情感陪伴AI使用率近半，在AI使用动机上，孤独感、情绪支持、社交需求是主要使用原因，娱乐也有一定需求。AI使用体验上，整体体验较好，特别是在隐私保护和个性化程度上评价较高。对自然交流和情感支持、隐私保护有较高需求，使用场景集中在情绪低落和睡前，对隐私保护和道德问题关注较多，付费意愿中等偏上、对价格较敏感、更期待买断制，选购产品时看重对话质量、隐私保护、个性化定制和功能丰富性。

综上，这一类用户是情感陪伴AI的核心用户，以年轻的在校学生为主，由孤独感、抑郁度产生的情绪支持需求驱动了对情感陪伴AI的较高使用率和付费意愿。他们对隐私保护和个性化定制的需求很高，在睡前和情绪低落时使用较多，但是消费能力一般、对价格较敏感。

**第二类群体（165人，占比44%）：**

人群特征上，整体年龄偏小，消费水平较低、女性比例偏低，且孤独感与抑郁程度均偏低，以在校本科学生为主。情感陪伴AI的使用率略低，约占四成，他们的使用动机主要集中在情绪支持，同时也存在一定的孤独感缓解和娱乐需求。AI使用体验整体较好，尤其在自然交流方面获得较高评价，但在隐私保护和个性化定制上则略显不足。功能需求上，他们对情感支持和隐私保护有明确要求，而对个性化定制的依赖相对较低。使用场景主要集中在睡前及娱乐时，同时也有一定的情绪支持需求。付费意愿较弱，对价格非常敏感，更倾向于订阅或低价买断制。

综上，该群体为情感陪伴AI的潜在用户，主要由年轻学生构成，情感陪伴AI使用率较低，其中的使用者以消遣娱乐为主，同时对情绪支持有一定需求，但整体消费能力有限、支付意愿较弱，更注重实用性与隐私保护。

**第三类群体（92人，占比24%）：**

这一群体的用户年龄相对较大，消费能力较强，同时孤独感与抑郁程度均处于较低水平，主要以全职工作者为主，女性比例较高。情感陪伴AI的使用动机较为弱化，用户主要因轻微的情绪支持和缓解孤独而使用，但整体需求并不强烈。AI使用体验的评价一般，对隐私保护和个性化体验的要求不如核心用户明显。功能需求方面，他们对隐私保护较为看重，但对情感支持和自然交流的依赖度较低。使用场景上，除了常见的睡前和情绪低落时使用外，还会在工作、旅行等多种情境下尝试。付费意愿较高，但对价格保持理性，倾向于一次性购买。

总体来看，该群体为成熟职场人士，尽管对情感陪伴需求不突出，但凭借较高的消费能力，他们更关注产品的整体品质、隐私安全及丰富功能，同时更青睐具有较好品牌口碑的产品。

3.营销建议

针对用户画像，企业可以定制差异化的营销策略，提升情感陪伴AI在各细分市场的核心竞争力。

对于第一类群体，在产品研发中着重提升情感支持和自然交流的能力，加快技术迭代；同时个性化定制，制定严格的隐私保护协议。营销推广可以重点围绕校园渠道展开，通过赞助校内活动、开展线上互动体验等方式，让以学生为主的用户切实感受到产品在缓解孤独、提供情感支持等方面的优势；同时通过用户故事、专家背书，树立产品在隐私保护和定制服务等方面的品牌形象，吸引核心用户的持续关注和深度体验。

对于第二类群体，由于情感支持需求较低、对价格很敏感，产品在设计时应注重实用性和性价比，突出基础但实用的情感支持功能，适当简化个性化定制选项，以降低产品开发和运营成本。推广策略上，企业可通过灵活的订阅模式或低价买断方案降低用户的尝试门槛，并通过社交媒体和校园大使的推荐来提升品牌曝光度和产品信任感。

对于第三类群体，作为成熟职场人士，对情感陪伴AI需求并不突出，但消费能力较强，所以产品设计应强化品牌的高端形象；营销推广策略则应聚焦于高端市场，借助专业媒体报道、职场社交平台及企业合作等渠道，将产品定位为高品质、值得信赖的情感陪伴工具。与此同时，提供买断制和个性化定制服务，满足他们长期使用和个性需求，并通过整合线上线下资源构建高端用户社群，进一步提升品牌忠诚度和市场影响力。