

# Dognition 公司业务数据分析报告

## 一、 分析背景

Dognition 是一家指导用户对宠物狗进行特定测试的公司，它能够教用户如何与他们的宠物狗建立更深层次的联系，并给用户一个前所未有的视角来了解他们的宠物狗的个性和能力。该公司提供了 20 个由科学家、训练师和行为专家制作的关于如何与宠物狗进行有趣的互动游戏的详细说明和操作视频。当用户与宠物狗完成这 20 个互动游戏后，Dognition 会提供一份 10 到 15 页的关于用户宠物狗的独特个性维度的报告给用户，该报告为用户提供个性化的见解，了解用户的宠物狗用于与世界互动的认知策略，并深入分析宠物狗在每个游戏中与其他狗相比的表现。同时用户还会被告知其宠物狗的性格特征，这是九种性格特征之一，比如“社交名流”、“爱因斯坦”或“特立独行者”。

## 二、 分析目的

Dognition 公司的主要目标之一就是能够从尽可能多的不同种类的狗身上收集尽可能多的数据。因此，该业务目的是帮助 Dognition 公司弄清楚他们可以实现哪些业务变更来增加用户在他们网站上完成的测试数量。

## 三、 分析思路

### （一）数据描述

该业务使用的数据集由 Dognition 公司提供，该数据集储存于 EXCEL 表格中，包含两个数据表，数据表 1 记录了每位用户的每只宠物狗进行测试的情况，数据量为 177481 行，具有 29 个字段（表 1）；数据表 2 记录了各用户所在地区与 UTM 时间的时间差，数据量为 42521 行，具有 2 个字段（表 2）。

表 1. 测试记录信息表

字段名	描述
Created_at	用户输入宠物狗的测试结果时的时间戳
Updated_at	用户上次更新宠物狗的测试结果的时间戳
User_ID	用户的唯一 ID
Dog_ID	宠物狗的唯一 ID
Test_name	Dognition 测试的名称
Subcategory_name	测试所属的子类别名称
Gender	宠物狗的性别
Birthday	宠物狗的出生年份

Breed	品种名称
Breed_Type	犬种类型(纯种、混合种、杂交种)
Breed_Group	犬种所属的群体(放牧、猎犬、非运动、运动、梗、玩具、工作)
Weight	宠物狗的体重(磅)
Dog_Fixed	标志狗只是否绝育(1=是/0=否)
DNA_Testcd	标志是否测试宠物狗的DNA(1=是/0=否)
Dimension	Dognition 公司的 9 大个性特征之一(Ace、Charmer、Einstein、Expert、Maverick、Protodog、Renaissance-dog、Socialite、Stargazer)
Sign_in_Count	用户登录其 Dognition 账户的次数
Max_Dogs	与用户关联的宠物狗的数量
Subscribed	标记用户是否具有付费订阅(1=是/0=否)
City	用户所在城市
State	用户所在州
Zip	用户所在地区的邮政编码
Country	用户所在国家
Exclude	标志该记录是否应被排除(1=排除, 0=不一定排除)
Free_Start_User	通过免费获得前 4 款游戏来指示用户是否获得了“免费开始”的标志(1=免费开始, 0=非免费开始)
Last_Active_At	用户在其 Dognition 账户中最近一次活动的时间戳
Membership_Type	订阅的类型(1=初始 20 款游戏的认知评估, 2=年度, 3=月度, 4=免费, 5=用户选择的订阅类型)
Rating	用户对于宠物狗测试结果惊讶程度(用户可以选择 1(不奇怪)到 9(非常奇怪)之间的任何数字)
Rank_by_UserID	用户完成的每个测试的时间顺序
Rank_by_DogID	宠物狗完成的每项测试的时间顺序

**表 2. 时区校正表**

字段名	描述
Zip	用户所在地区的邮政编码
Diff from UTC	用户所在地区的时间与 UTC 时间的差值

## （二）分析框架

根据分析目的，建立 S.M.A.R.T 目标，对分析目的进行精细化，精细化目标为：根据每个用户和宠物狗分别的平均完成测试数量和总完成测试数量，进行探索性分析，确定能够影响测试数量的特定的特征。

根据精细化目标，建立结构金字塔分析框架（图 1），该业务分析中，因变量为每位用户的总完成测试数量、每位用户的平均完成测试数量、每只宠物狗的总完成测试数量、每只宠物狗的平均完成测试数量。自变量主要分为 3 大类，第一类为宠物狗的特征，主要包括了宠物狗的性别、年龄、个性特征、品种（品种类型、品种群体）、体重、是否阉割、是否经过 DNA 测试。第二类为用户的特征，主要包括用户登录其账户次数、拥有宠物狗数量、所属成员类型、是否付费订阅。第三类为测试环境特征，包括进行测试所在时间（年、月、日、时）、进行测试所在地理位置（国家、州）。

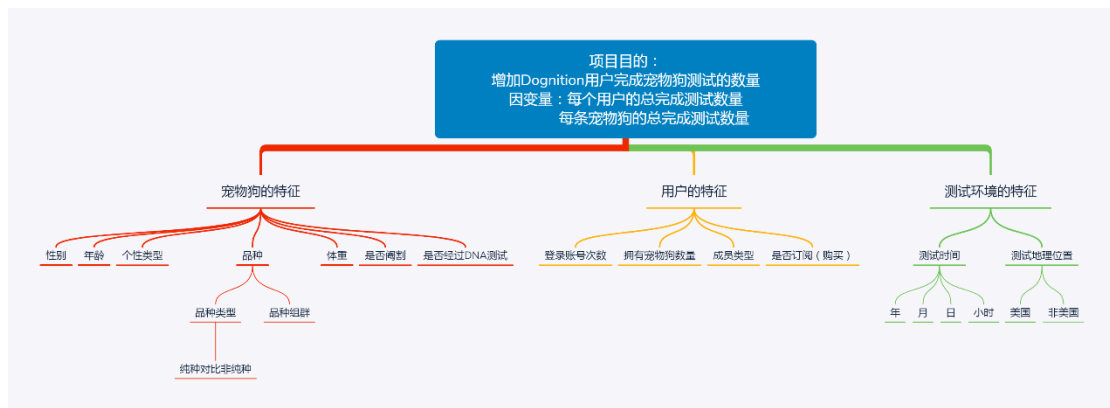


图 1. 结构金字塔分析框架

## 四、 分析内容

根据分析框架，将探索性分析内容分为 3 部分，包括宠物狗特征、用户特征及测试环境特征。

### （一）宠物狗特征

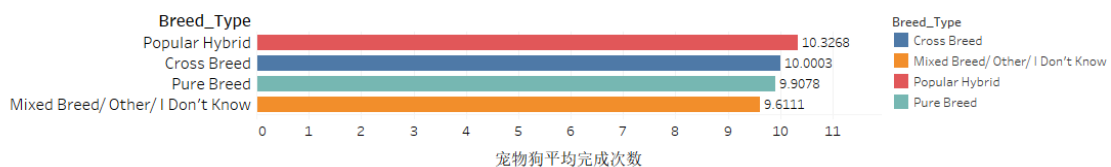


图 2. 宠物狗品种类型与平均完成测试数量的关系

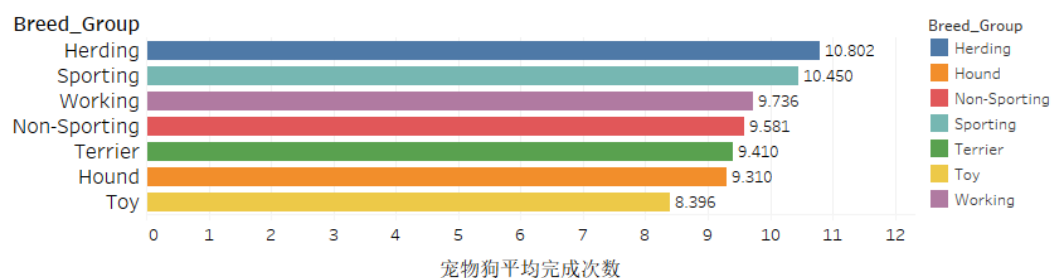


图 3. 宠物狗品种群体与平均完成测试数量的关系

对宠物狗的品种类型进行分析，由图 2 可知，受欢迎的混合犬类型的宠物狗平均完成次数最高，而品种类型为未知的宠物狗平均完成次数最低，这可能是由于受欢迎的混合犬宠物狗品种类型受主流大众的喜爱，其价格较高，用户会花更多心思在该种类宠物狗身上，更希望能够了解自己小狗的特性。而对于品种类型为未知的宠物犬，多为普通犬类，用户对于了解其个性的兴趣较低。

图 3 展示的是宠物狗品种群体与平均完成测试数量的关系，宠物狗被分成 7 种群体，包括放牧（Herding）、体育（Sporting）、工作（Working）、非体育（Non-Sporting）、梗犬（Terrier）、猎犬（Hound）、玩具（Toy）。由图可知，放牧犬的平均完成次数最多，为 10.802 次，体育犬次之，为 10.45 次，工作、非体育、梗犬及猎犬平均完成次数在 9.5 左右，而玩具犬的平均完成次数最少，只有 8.396 次。

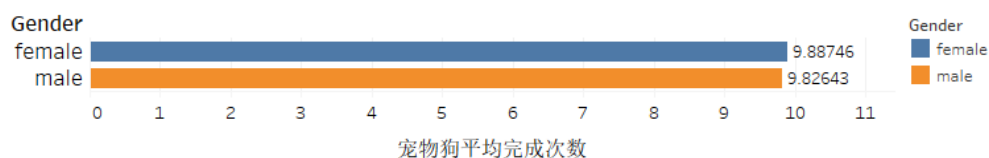


图 4. 宠物狗性别与平均完成测试数量的关系

图 4 分析了宠物狗性别与平均完成测试数量的关系，由图可知，宠物狗的性别对用户对于是否对宠物狗进行测试的决定影响不大。

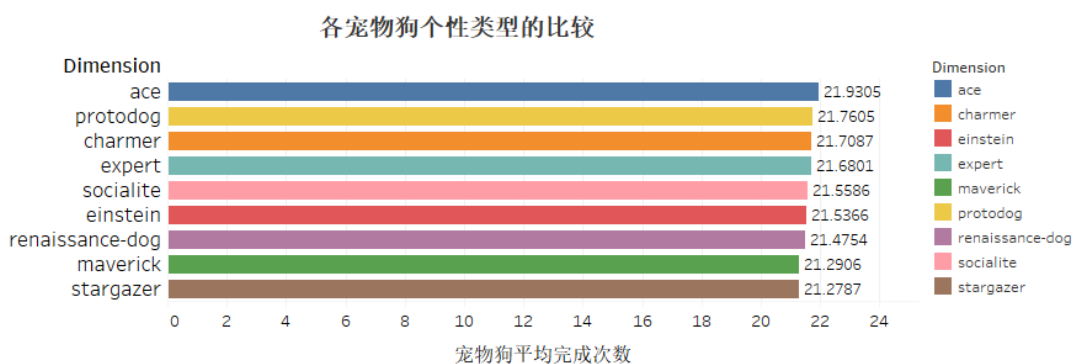


图 5. 宠物狗个性类型与平均完成测试数量的关系

若宠物狗通过 20 个测试，Dognition 会提供一份报告给用户，让其更好的了解宠物狗所对应的个性类型。主要分为 9 种，包括佼佼者（Ace）、模范（Protodog）、魔术师（Charmer）、专家（Expert）、社交名流（Socialite）、爱因斯坦（Einstein）、文艺复兴（Renaissance-dog）、特立独行（Maverick）以及占星师（Stargazer）。由图 5 可得，各个性类型的宠物犬的平均测试完成数量差别不大，主要在 21 次左右。



图 6. 宠物狗是否阉割与平均完成测试数量的关系

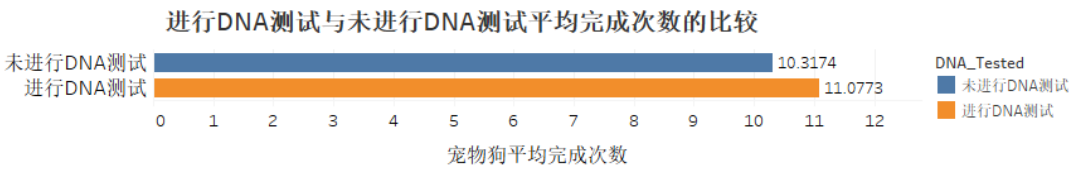


图 7. 宠物狗是否进行 DNA 测试与平均完成测试数量的关系

图 6 和图 7 分别分析了宠物狗是否阉割和宠物狗是否进行 DNA 测试与平均完成测试数量的关系，这两个因素能够很好的反应用户的性格，由图可以看出，进行阉割的宠物狗的平均完成测试数量较高于未阉割的宠物狗，这可能由于阉割能够有效的减少宠物狗的患病几率，及宠物狗的攻击性，这能够表明了用户对于其宠物狗的关注程度，所以可以推断用户本身对于更多地了解他们的宠物狗感兴趣，因此这些对宠物狗进行阉割的用户将比其他用户更有可能完成 Dognition 测试组，由于测试将提供他们关于宠物狗的更多见解。而同样的，进行 DNA 测试的宠物狗的平均完成测试数量大于未进行 DNA 测试的宠物狗，这可能是由于进行 DNA 测试能够使用户了解其宠物犬是否为纯种犬，同时能让用户了解其宠物犬的源地，这同样表示了用户对于其宠物狗有较高的感兴趣程度。

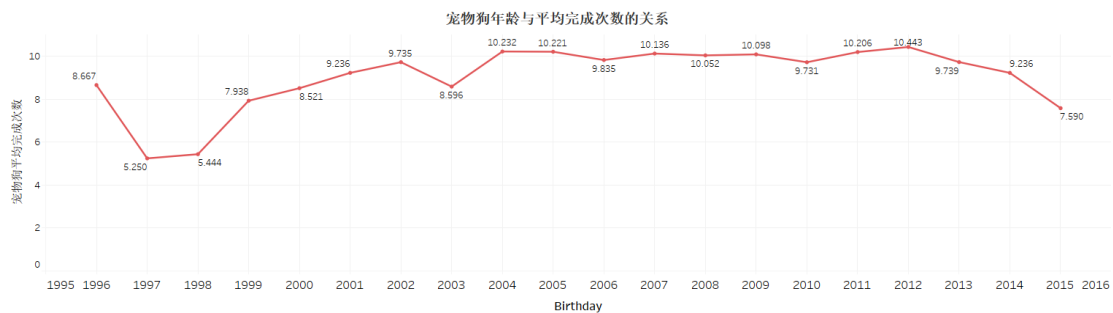


图 8. 宠物狗的年龄与平均完成测试数量的关系

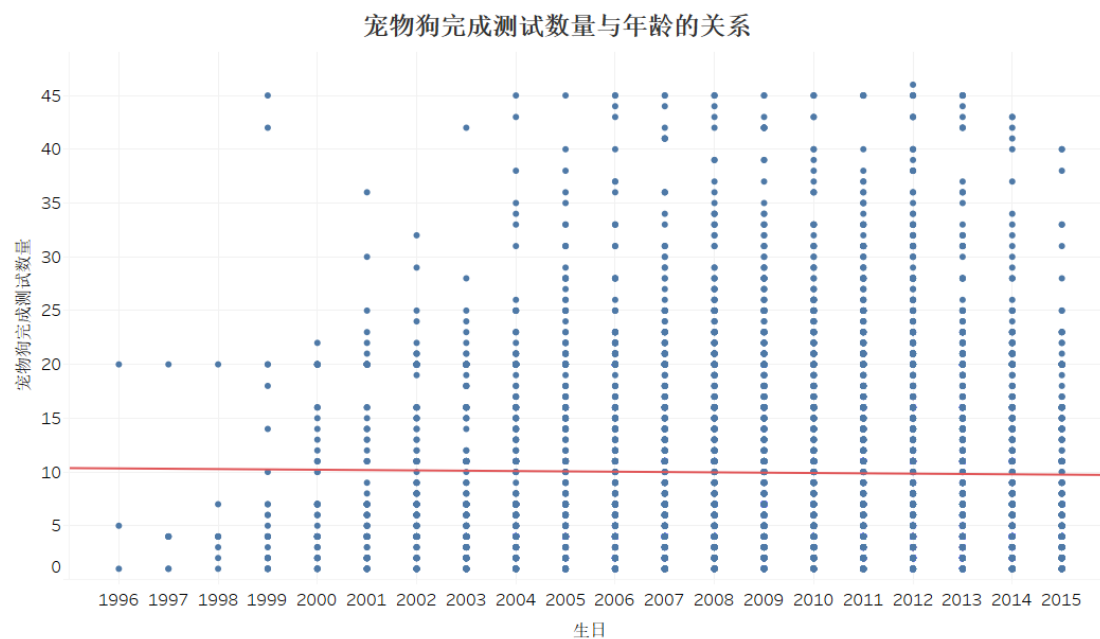


图 9. 宠物狗的年龄与总完成测试数量的关系

关于宠物狗的年龄，图 8 分析了年龄与平均完成测试数量的关系，由图可知，平均完成测试数量在宠物狗的出生年份为 1997 年时达到最低，而在之后逐渐升高，在 2004 年份升高为 10.232，之后平均完成测试数量趋于稳定。由于年龄变量为连续变量，所以建立各宠物狗的年龄与总完成测试数量的散点图，并进行相关性分析，创建趋势线，由图 9 可知，宠物狗的年龄与总完成测试数量的趋势线接近于一条水平直线，这表明宠物狗的年龄特征与用户对宠物狗进行测试数量的决定几乎无关。而两张图在 2004 年份之前表现内容有一些不同的原因可能在于，在 2004 年份之前，宠物狗的数量较少，所以平均完成测试数量起伏较大。

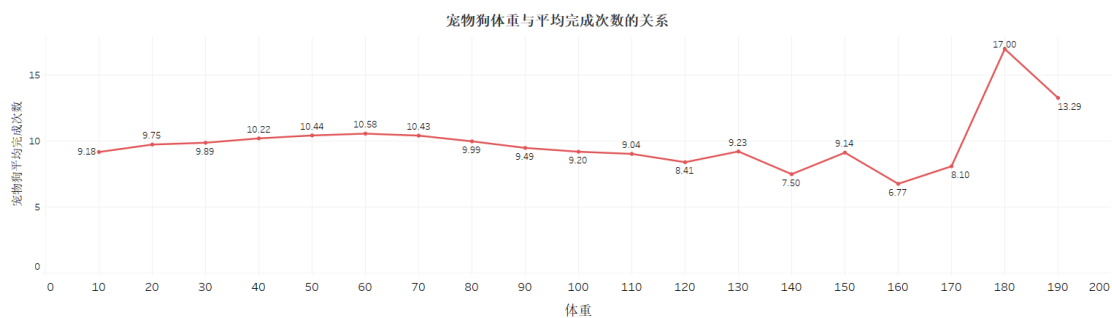


图 10. 宠物狗的体重与平均完成测试数量的关系

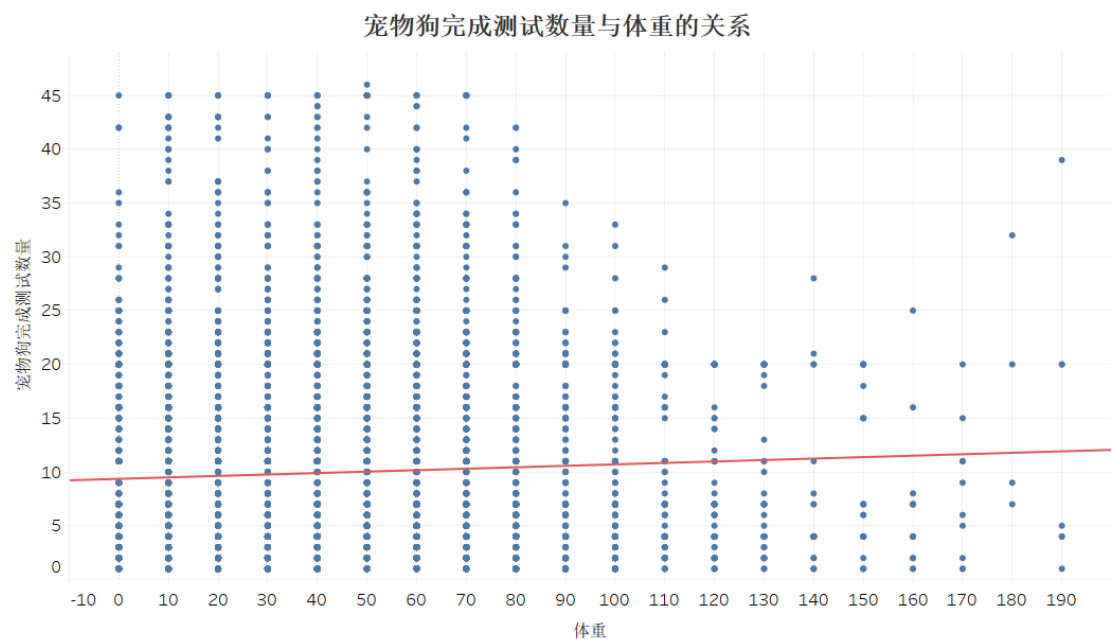


图 11. 宠物狗的体重与总完成测试数量的关系

同样的，宠物狗的体重特征也是连续变量，图 10 和图 11 分析了宠物狗的体重与平均完成测试数量和总完成测试数量的关系，由图 10 可知，宠物狗的平均完成测试数量随着体重的增加起伏不大，从 10 磅到 130 磅，平均完成测试数量都为 10 左右，而在 140 磅，平均完成测试数量最低，为 6.77，而 180 磅宠物狗的平均完成测试数量达到最高，为 17。然而通过图 11，可以观察到宠物狗的体重与总完成测试数量的趋势线坡度极缓，这表示宠物狗的体重与用户对宠物狗进行测试数量的决定关系不大。而两张图在 130 磅后表现内容有一些不同的原因可能在于，在 130 磅以后，宠物狗的数量较少，所以平均完成测试数量起伏较大。

## （二）用户特征

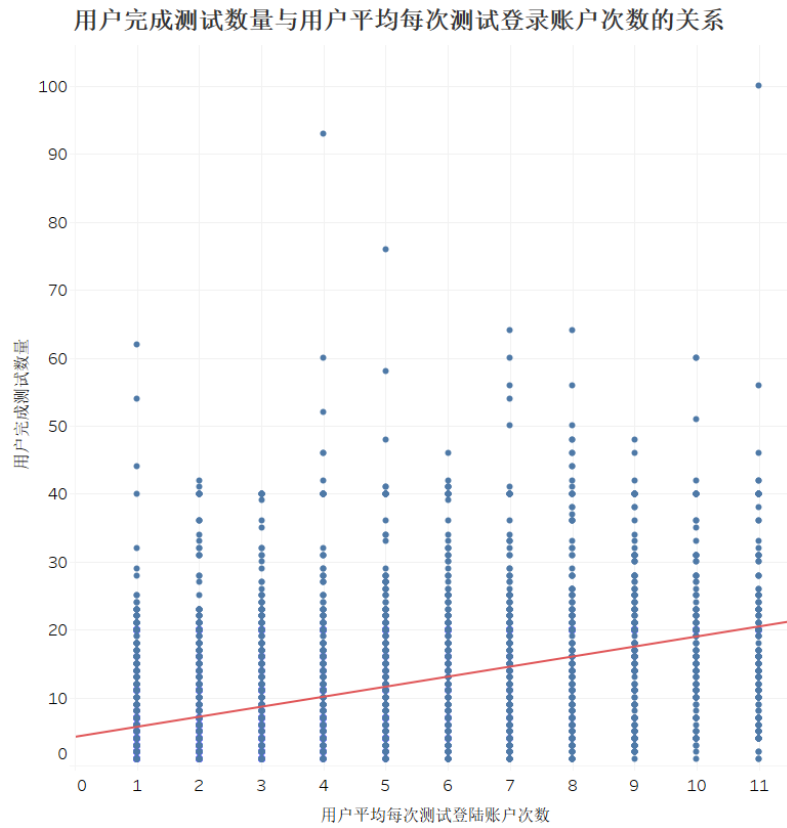


图 12. 用户平均每次测试登录账户次数与总完成测试数量的关系

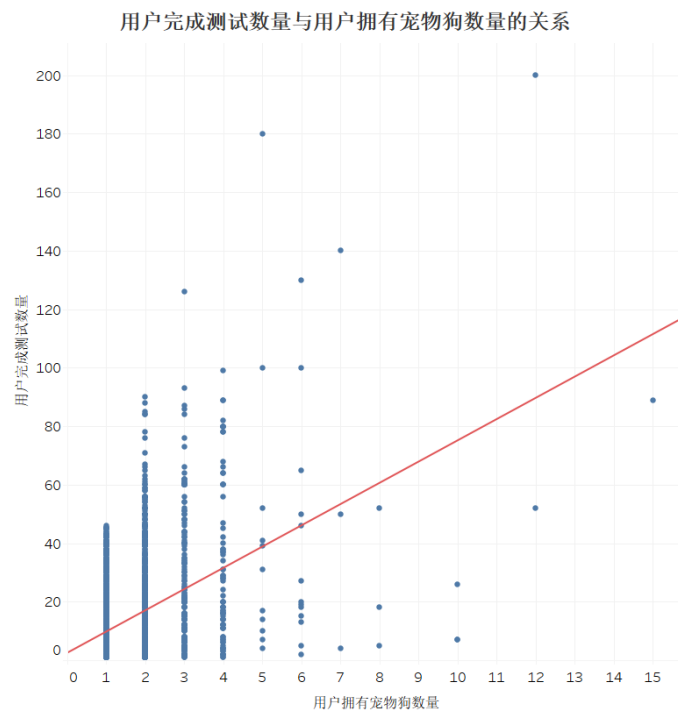
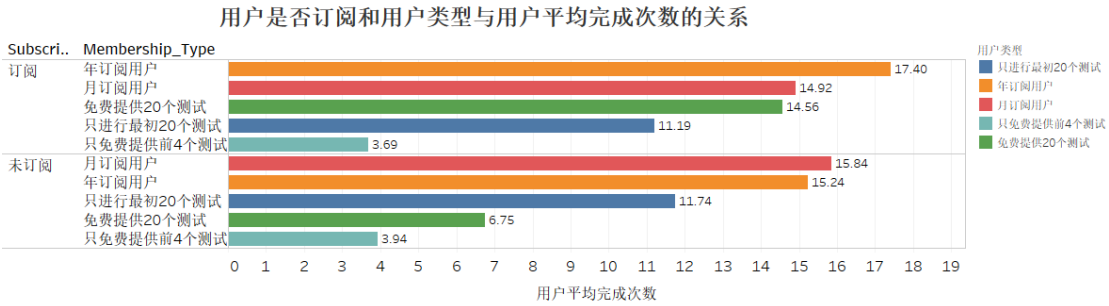


图 13. 用户拥有宠物狗数量与总完成测试数量的关系



图 12 分析了用户平均每次测试登录账户次数与总完成测试数量的关系，由于用户平均每次登录账户次数为连续变量，所以利用散点图对其进行相关性分析，并创建两个变量的趋势线。由图 12 的趋势线可知，用户平均每次测试登录账户次数与总完成测试数量具有正相关性。这表明，用户平均登录账户次数越多，用户的总完成测试数量也会越高，这可能是由于用户越频繁的登录账户，代表其对 Dognition 公司所提供的测试更加重视及更有兴趣。

图 13 分析了用户拥有宠物狗数量与总完成测试数量的关系，选择相关性分析方法，建立两个变量的散点图，并创建趋势线。由图 13 的趋势线可知，用户拥有宠物狗数量与总完成测试数量具有极强的正相关性。这同样也表明，用户拥有宠物狗数量越多，用户的总完成测试数量也会越高，这可能是因为，豢养较多宠物狗的用户对宠物狗关心及感兴趣程度会大于豢养较少宠物狗的用户，所以豢养宠物狗数量多的用户对 Dognition 公司所提供的测试更加重视及更有兴趣。



**图 14. 用户是否订阅和用户类型与平均完成测试数量的关系**

用户特征还包括了两个特征，一是用户是否付费订阅了 Dognition 的测试，第二个特征为用户所对应的类型，具体有 5 种类型，包括年订阅用户、月订阅用户、免费提供 20 个测试用户、只进行前 20 个测试的用户，以及免费提供前 4 个测试的用户。图 14 分析了用户是否订阅和用户类型与平均完成测试数量的关系，采用对比分析方法，比较了用户是否订阅对不同类型用户平均完成测试数量的影响。从图 14 可以看出，订阅用户的平均完成测试数量明显高于未订阅用户，并且对于年订阅用户而言，选择订阅的用户平均完成测试数量要高于未选择订阅的用户，然而对于月订阅用户而言，选择订阅的用户平均完成测试数量要低于未选择订阅的用户。这可能是由于大多数用户完成 20 个测试的时间范围要大于 1 个月，所以当月订阅用户处于未订阅状态时，其仍然能够继续进行 20 个测试中未完成的部分，而对于订阅状态的月订阅用户来说，他们还正处于刚开始进

行测试的阶段，所以平均完成测试数量会较少于未订阅用户。

从图 14 上还可以发现，免费提供 20 个测试的订阅用户与未订阅用户的平均完成测试数量存在巨大的差异，平均完成测试数量的差值接近于 8 次，这主要原因很可能是免费提供 20 个测试给用户后，在用户初步尝试测试后，若用户对参与测试和了解其宠物狗的个性产生了兴趣，用户便会选择进行订阅，若用户未产生兴趣，则很可能就会终止测试，并选择不订阅。所以订阅状态的用户平均完成测试的数量会远高于未订阅状态的用户。

（三）测试环境特征

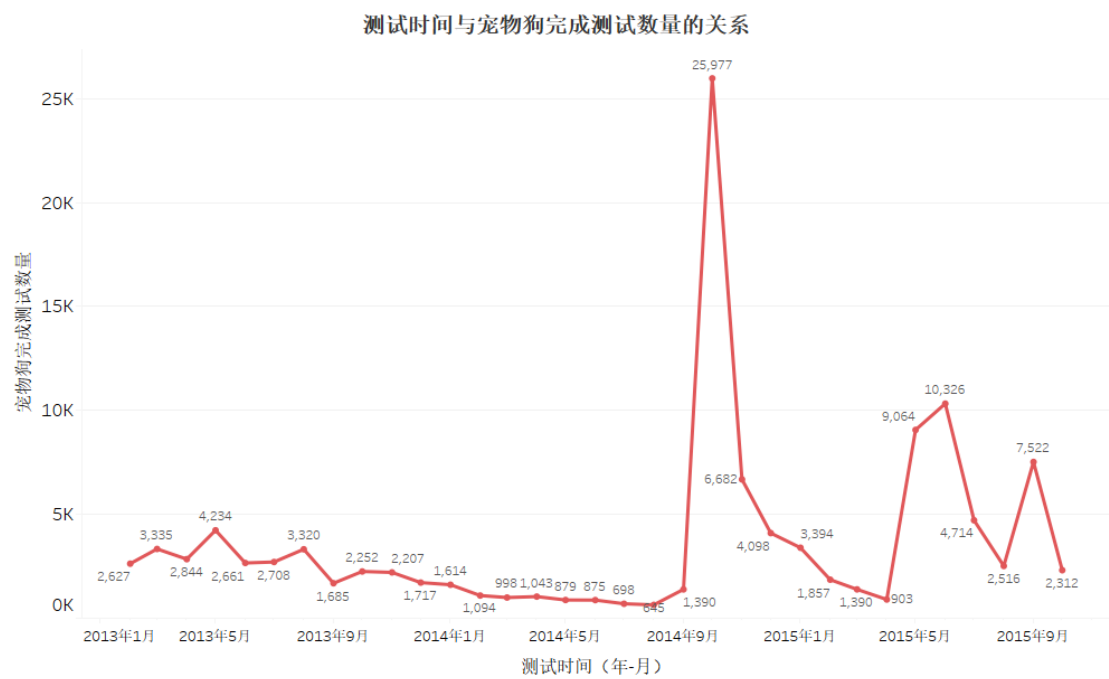


图 15. 宠物狗完成测试时间与总完成测试数量的关系

关于测试环境特征，主要分为两个部分，第一个部分为测试完成所在的具体时间，由于时间是连续变量，所以选择了趋势分析方法，建立折线图，这样能够有效的了解宠物狗完成测试数量的变化趋势。图 15 分析了用户完成测试间距（年-月）与总完成测试数量的关系，可以明显的看出从 2013 年 1 月份开始，宠物狗总完成测试数量有轻微波动，在 2500 左右，然而从 2013 年 10 月开始，总完成测试数量开始逐渐下降，并在 2014 年 8 月达到最低，为 698 次。但在 2014 年 10 月时，宠物狗的总完成测试数量发生巨大变化，迅速的攀升到了 25977 次。同样的在 2015 年 5 月和 6 月，宠物狗的总完成测试数量达到了较大的增长，达到 9064 次和 10326 次。

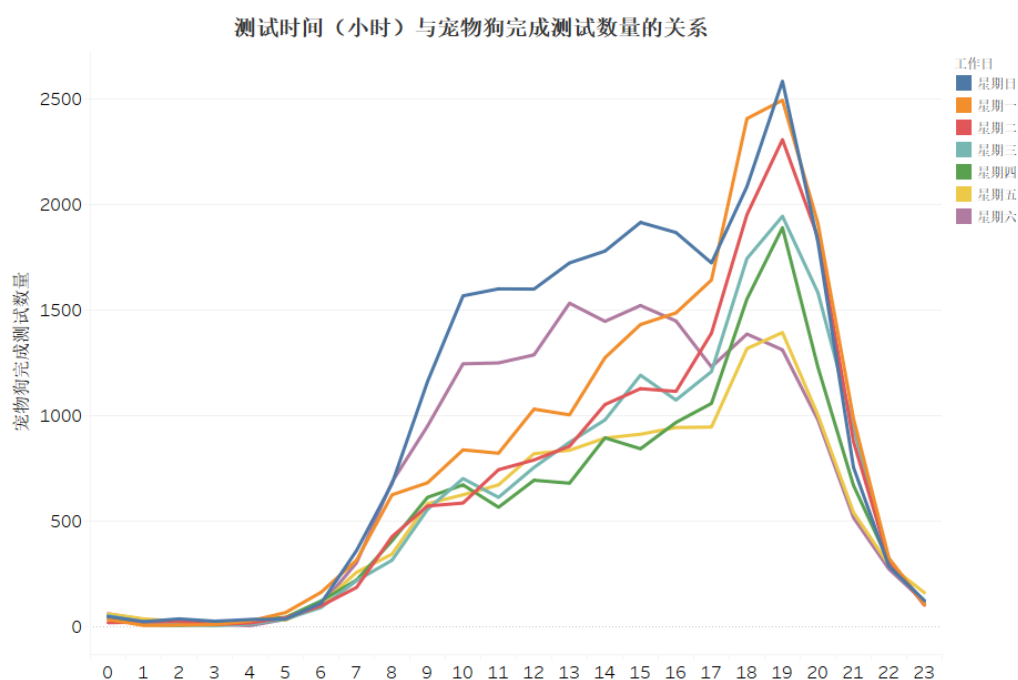


图 16. 宠物狗完成测试时间（小时）与总完成测试数量的关系

图 16 分析了宠物狗在一天中完成测试的时间与总完成测试数量的关系，并对比了一周内不同天数的完成情况，采用了对比分析和趋势分析的方法，建立折线图，并以不同的颜色的折线代表不同的天数。由图上可以看出，宠物狗总完成测试次数在星期日最多，而在星期五最少，并且宠物狗完成测试的时间主要集中在 10 时到 21 时这一范围。并且，对于星期一、二、三、四、五、日而言，从一天 0 时开始，完成测试的数量在逐渐上升，并在 19 时的时候，宠物狗总完成测试数量会达到一天中的顶峰。但在星期六，情况却有所不同，宠物狗完成测试的数目从 0 时可是逐渐上升，但到了 10 时至 20 时，宠物狗完成测试的数量变化不大，并没有特别的峰值点。

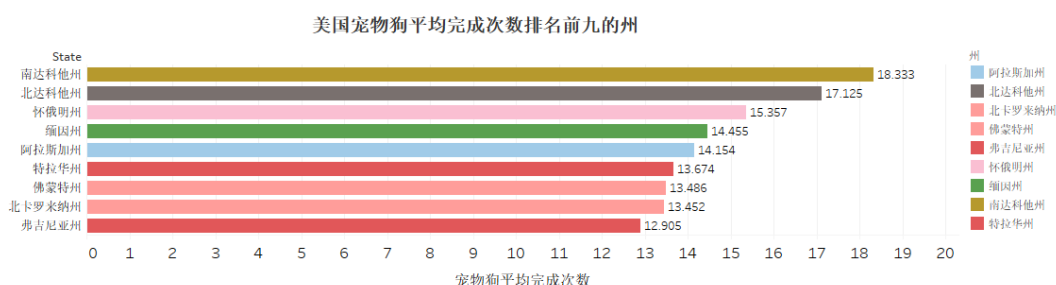


图 17. 宠物狗所在地区与平均完成测试数量的关系

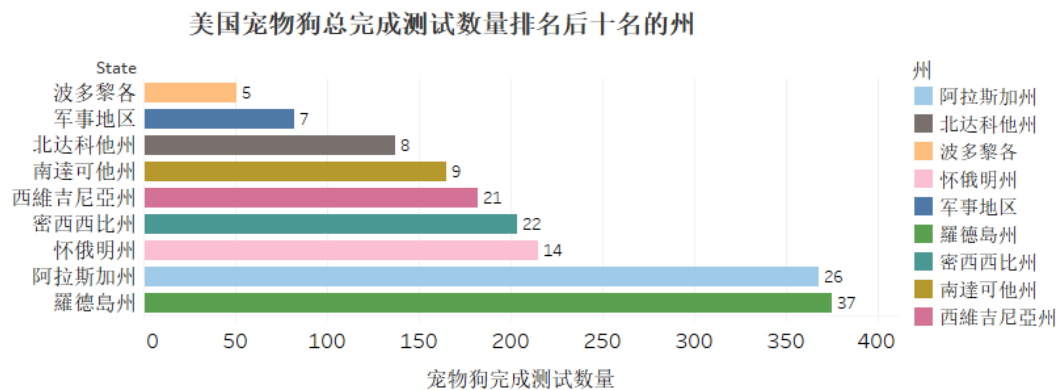


图 18. 宠物狗所在地区与总完成测试数量的关系

测试环境特征的第二个部分为宠物狗进行测试的所在地，所在地区不同，对于用户选择对宠物狗进行测试的决定影响也不同。由于 Dognition 公司位于美国，其主要面向对象为美国用户，所以选择对美国各州进行对比分析。图 17 分析了宠物狗所在地区与平均完成测试数量的关系，展示了美国宠物狗平均完成测试数量排名前九的州，从图上可知，南达科他州的平均完成测试数量最高，达到了 18.33 次，而北达科他州较为次之，为 17.125 次。图 18 分析了宠物狗所在地区与总完成测试数量的关系，展示了美国宠物狗总完成测试数量排名最后十名的州，从图中可知，波多黎各的宠物狗完成测试数量最少，只有 5 次。并且从图 17 和图 18 中可以发现，虽然北达科他州、南达科他州、怀俄明州以及阿拉斯加州的平均完成测试数量位于美国各州的前十位，但其总完成测试数量却位于美国各州的后十位。

## 五、 结论与建议

### （一）结论

通过对数据分成了三个部分进行分析，结论主要包含三个方面：

#### （1） 宠物狗特征

该业务分析中分析的宠物狗特征有 8 个，为性别、年龄、体重、品种类型、品种组群、是否阉割、是否进行 DNA 测试，以及通过 Dognition 分析得到的个性类型。通过分析得知，不同性别、个性类型的宠物狗平均完成测试数量差别不大，并且宠物狗总完成测试数量与宠物狗的体重和年龄相关性不大，即宠物狗总完成测试数量并不会随着宠物狗的体重或年龄的变化而变化。

但在 4 个品种类型中，受欢迎的混合犬品种类型的宠物平均完成测试数量最高，而未知的品种类型的宠物平均完成测试数量最低。在 7 个品种组群中，放

牧和体育的品种组群具有较高的平均完成测试数量，工作、非体育、梗犬、猎犬的平均完成测试数量相比前两个品种组群较低一些，排在最后的是玩具品种组群，它与放牧犬的平均完成测试数量的差值达到了 2 次。

同样的，宠物狗是否阉割及是否进行 DNA 测试的平均完成测试数量存在差异，通过分析得出，进行阉割的宠物狗的平均完成测试数量要高于未进行阉割的宠物狗，而进行 DNA 测试的宠物狗的平均完成测试数量也要高于未进行 DNA 测试的。

## （2） 用户特征

该业务数据分析中分析的用户特征具有 4 个，为用户登录账户次数，拥有宠物狗数量、用户类型和用户是否付费订阅。通过相关性分析方法，建立散点图，创建趋势线，分析得出用户平均每次登录账户次数和用户拥有宠物狗数量都与用户完成测试数量有较强的正相关性，也就意味着，用户完成测试数量能够随着用户平均每次登录账户次数的增加而增加，同样的，用户完成测试数量也能够随着用户拥有宠物狗数量的增加而增加。

通过选择对比分析的方法，对是否付费订阅的用户的平均完成测试数量进行对比，并结合不同的用户类型，发现订阅用户的平均完成测试数量明显高于未订阅用户，并且付费订阅的年订阅类型用户具有最高的平均完成测试数量，而对于免费提供 20 个测试的订阅用户与未订阅用户的平均完成测试数量存在巨大的差异。这主要原因很可能是免费提供 20 个测试给用户后，在用户初步尝试测试后，若用户对参与测试和了解其宠物狗的个性产生了兴趣，用户便会选择进行订阅，若用户未产生兴趣，则很可能就会终止测试，并选择不订阅。所以订阅状态的用户平均完成测试的数量会远高于未订阅状态的用户。这也表明免费提供 20 个测试给用户的业务方案具有较好的成果。

## （3） 测试环境特征

该业务数据分析中分析的测试环境特征为 2 个部分，第一部分为测试时间，第二部分为测试地点。选择使用趋势分析的方法，建立了宠物狗完成测试数量随着测试时间变化的趋势折线图，并通过分析得出，在 2014 年 10 月和 2015 年 6 月，宠物狗完成测试的数量有了大幅度的提升。同样的，采用对比分析和趋势分析的方法，建立折线图，并以不同的颜色的折线代表不同的天数，分析了宠

物狗在一天中完成测试的时间与总完成测试数量的关系,并对比了一周内不同天数的完成情况。由此发现宠物狗总完成测试次数在星期日最多,而在星期五最少,并且除星期六以外,其他天数的宠物狗完成测试数量都在 19 时达到了一天的顶峰,这说明用户倾向于在星期日,并且 19 时左右对宠物狗进行测试。

第二部分为测试地点,通过对比分析的方法,分析了美国各州的平均完成测试数量和总完成测试的差异。得知虽然北达科他州、南达科他州、怀俄明州以及阿拉斯加州的平均完成测试数量位于美国各州的前十位,但其总完成测试数量却位于美国各州的后十位。

## (二) 建议

通过分析,主要有以下几点建议:

### (1) 增加优惠活动对特定宠物狗类型的用户

通过分析我们可以得知,用户对于未知品种类型、玩具犬组群、未阉割及未进行 DNA 测试的宠物狗进行 Dognition 测试的兴趣感不强,可开展特定业务针对拥有这些类型的宠物狗的用户,如对于这种类型的用户,将部分测试作为免费测试开放。

### (2) 在特定时间对用户进行宣传及通知

通过分析我们可以得知,用户倾向于在周日及 17 时至 19 时对宠物狗进行测试,Dognition 公司可以在该时间段对用户进行 Email 提醒,这样能够增加用户测试宠物狗的可能性。同样的,对于用户不倾向对宠物狗进行测试的时间,如早上 9 时或星期五这天,Dognition 公司可以在这些时间段开发免费测试活动,或一些奖励活动,这样能够吸引在这些时间段有时间的用户对宠物狗进行测试。

### (3) 对特定的地区进行大力宣传

通过分析我们可以得知,波多黎各、北达科他州、南达科他州、怀俄明州、阿拉斯加州等州的宠物狗完成测试数量远低于其他的州,所以 Dognition 公司需要在这些排名低的地区加大宣传力度,并给予这些地区的用户特定的活动优惠,如完成一定数量的测试可以有特殊的奖励等。这样能够吸引这些地区的用户对其宠物狗进行测试。