



西安交通大学
XI'AN JIAOTONG UNIVERSITY

CryptoKitties 数据分析报告



课程名称： 大数据技术基础

姓名： 李峥昊

学院： 管理学院

专业： 大数据管理与应用

学号： 2201111618

2023 年 2 月 17 日

一、 引言：区块链和以太坊

在现在的互联网的架构中，在进行交易时，由于我们需要一个平台来对信息进行整合并提供一定的担保，从而引发出了两个问题：一是信任问题，由于个人数据集中存储和管理在服务器上，由大型公司掌控，因此个人数据很容易被收集、分析和使用；并且由于大型公司控制了数据，从而它们有能力控制和监督网络上的内容。二是价值分配问题，互联网上有众多的使用者去分享他们的产品或服务，但是财富大多是由提供服务的公司积累的。以音乐行业为例，据调查数据显示，由于中间商多、价值链复杂、版权数据不透明，音乐创作者权益处于弱势地位，目前基于互联网流音乐的版税分配中，一首歌的版税分给唱片公司、操作系统和互联网播放平台的比例分别为 42%、30%、和 20%，而词、曲作者总共只能分到 8%[1]。

区块链（Block Chain）技术使用哈希指针代替普通指针，通过哈希函数的 collision resistance 和 hiding 两个性质使得区块链中的数据无法被篡改而不被发现。从而保证了数据的安全性和不可更改性。同时引入分布式记账和共识机制，使多台分散的主机共同维护一个网络。

以太坊（Ethereum）是区块链的一个应用。与比特币（Bitcoin）不同的是，以太坊通过一个状态树来记录所有有余额的用户数据，如余额、智能合约代码、存储数据等。通过交易树来记录以太坊网络上的所有交易信息，如交易的发起者、接受者、交易数据、费用等。通过收据树记录交易执行的结果信息，如交易是否执行成功、执行后的账户状态、智能合约产生的输出等。

此外，以太坊还支持智能合约的应用。智能合约是一种部署在区块链上的程序，用户通过调用智能合约以实现某些特定功能，并支付一定的 gas 作为调用计算资源的报酬。智能合约与前端界面共同构成了可供用户使用的去中心化应用（DApp, Distributed Application）。

二、 CryptoKitties 及其智能合约

CryptoKitties 是一款部署在以太坊区块链上的 DApp，是第一个大规模的以太坊 DApp，在 2017 年发布后受到了广泛的关注。它允许用户生成、购买、销售和交换 CryptoKitties，并且每只 CryptoKitty 都是一个独特的、不可复制的数字资产，具有一定的收藏价值。

CryptoKitties 的智能合约地址为 `0x06012c8cf97BEaD5deAe237070F9587f8E7A266d`，除了基因算法，其代码在以太坊上公开可见。CryptoKitties 智能合约的结构如图1所示。

其中，以下几个函数所产生的数据比较重要。这些函数主要集中在三个合约地址，Core: `0x0601_266d`，Sale Auction: `0xb169_7d8c` 和 Siring Auction: `0xc7af_7f26`。为简短起见，仅保留地址的开头四位和结尾四位。函数具体信息如表1所示。

- (1) createSaleAuction。创建一个计时拍卖，拍卖猫的所有权，价格随着时间线性变化。在用户成功创建拍卖后的同时，用户的猫将转移给 CryptoKitties 下的一个地址。用户可以随时无条件取消拍卖，取消拍卖时调用 cancelAuction 函数，将猫从 CryptoKitties 还给用户。
- (2) createSiringAuction。创建一个计时拍卖，拍卖猫的交配权，价格随着时间线性变化，用户可以随时无条件取消拍卖，取消拍卖时调用 cancelAuction 函数，将猫从 CryptoKitties 还给用户。如果交配成功，则卖方收回猫的所有权，买方猫怀孕并获得小猫的所有权。双方的猫共同进入休息期。
- (3) giveBirth。在交配以后，用户可以得到新的猫，其具体方式为：获取母亲和父亲的基因，通过基因算法（遗传和变异）来确定子代的基因，并得到新的小猫。在小猫出生后，将小猫的所有权移交给母亲的持有者上。
- (4) breedWithAuto。用户可以繁衍自己的猫，即自己的猫和自己的猫交配。

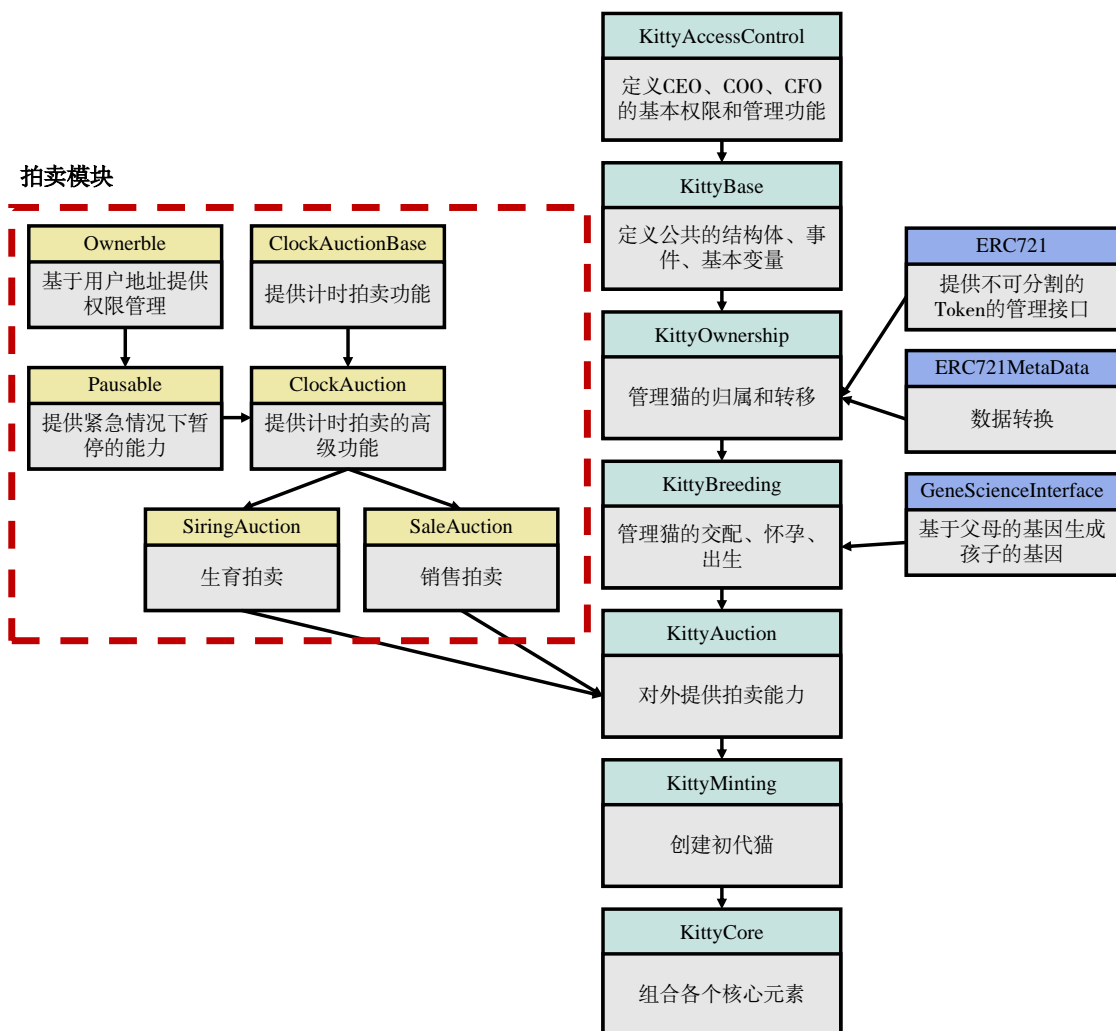


图 1: CryptoKitties 智能合约的结构图

- (5) transfer。用户可以将猫的所有权转移给另外一个人。
- (6) bidOnSiringAuction。用户可以对生育拍卖进行出价，竞拍成功后，用户获得一只已怀孕的猫，随后可以调用 giveBirth 获得新的小猫。
- (7) bid。用户可以对销售拍卖进行出价，竞拍成功后获得这只猫的所有权。

表 1: CryptoKitties 函数详情

name	contract	method	log	data
createSaleAuction	Core	0x3d7d3f5a	Transfer AuctionCreated	kittyId startingPrice endingPrice duration
createSiringAuction	Core	0x4ad8c938	Transfer AuctionCreated	kittyId startingPrice endingPrice duration
giveBirth	Core	0x88c2a0bf	Birth Transfer	matronId
breedWithAuto	Core	0xf7d8c883	Pregnant	matronId sireId
transfer	Core	0xa9059cbb	Transfer	address_to tokenId
bidOnSiringAuction	Core	0xed60ade6	Gene AuctionSuccessful Transfer Pregnant	sireId matronId
bid	SaleAuction	0x454a2ab3	AuctionSuccessful Transfer	tokenId
cancelAuction	SaleAuction	0x96b5a755	AuctionCancelled Transfer	tokenId
cancelAuction	SiringAuction	0x96b5a755	AuctionCancelled Transfer	tokenId
note: contract address				
Core:	0x06012c8cf97BEaD5deAe237070F9587f8E7A266d			
SaleAuction:	0xb1690C08E213a35Ed9bAb7B318DE14420FB57d8C			
SiringAuction:	0xC7af99Fe5513eB6710e6D5f44F9989dA40F27F26			

三、 数据获取

首先尝试从 sql 的 CryptoKitties 数据库获取数据，其代码如 listing 1 所示。图 2 中显示了 saleAction 函数的调用频次。可以看到，数据库存在较为严重的数据缺失。在 2018 年 1 月到 2018 年 12 月，以及 2021 年 5 月到 2022 年 2 月均没有数据。

Listing 1: 统计 saleAuction 调用频次

```
1 SELECT
2   day, COUNT(*)
3 FROM
4   dapp_cryptokitties.cryptokitties_0601_266d_function_createsaleauction
5 GROUP BY
6   day
```

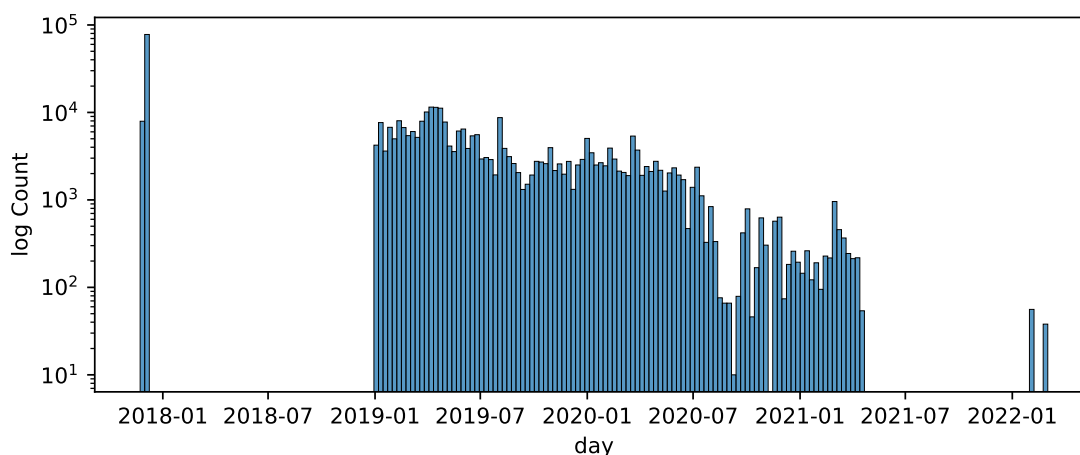


图 2: saleAction 函数调用频次

因此更换思路，从 hive 数据库获取需要的数据。数据获取过程分为两个部分：

- (1) 从 transactions 数据库获取地址为 Core、SaleAuction 的交易数据，并通过解析 input 来得到函数的输入。
- (2) 从 logs 数据库获取地址为 Core 的 log 数据，通过解析 data 来获得 log 的输入。

由于按照日期等于 “%Y-%m-%d” 检索速度较快，且如果一次性检索过多日期会触发 100,000 的数据量上限，因此采取网络自动化的方法，按日期检索。其 SQL 语句如 listing 2 所示。最后将分日期获取的 1865 个 csv 文件合并为一个文件，就获得了合约地址下的所有交易。最终生成了三个 csv 文件，分别为 Core.csv，SaleAuction.csv 和 log.csv。

Listing 2: 获取智能合约地址下的所有数据

```
1 SELECT
2   *
3 FROM
4   nifi_ethereum.transactions
```

```
5 WHERE
6   to_address = "0x06012c8cf97bead5deae237070f9587f8e7a266d"
7   AND day = "%Y-%m-%d"
```

四、 研究主题

1. CryptoKitties 的市场活跃度

市场活跃度，即市场中产品交易的数量和频率，是衡量市场或产品需求和流通状况的重要指标。市场活跃度高意味着需求和积极性高，说明市场或产品受欢迎，价值和发展潜力较高。

图3-图7展示了 2017 年 11 月至 2023 年 1 月的市场活跃度相关数据。其中横轴表示时间，14 天为一个单位；纵轴表示函数调用频次，以对数坐标呈现。

图3和图4展示了销售拍卖的相关数据。可以看到，从 2017 年至 2023 年，市场上的交易数量总体上呈现下降趋势，2018 年一整年，拍卖数量和成交数量都在每 14 天 1 万左右；而到了 2022 年下半年，拍卖数量和成交数量下降至每 14 天 120-150。

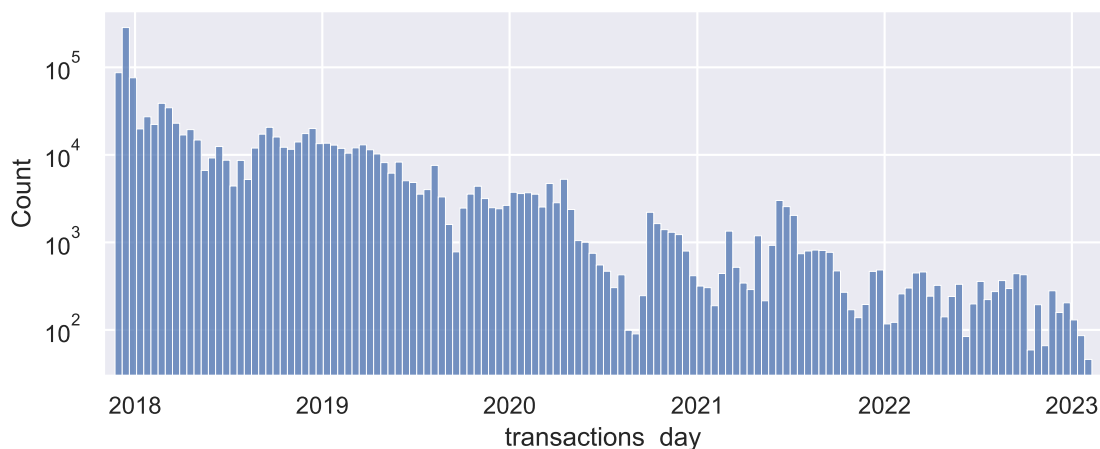


图 3: 创建销售拍卖的频次

图5和图6展示了生育拍卖的相关数据。可以看到，从 2017 年至 2022 年，拍卖数量和成交数量在整体上呈现下降趋势，逐步从 2018 年的每 14 天 1600 次下降至 2022 年的每 14 天 10 次。在 2022 年以后，几乎不再有人购买猫的交配权。但是拍卖次数在短暂的归零以后，又恢复到了每 14 天 10 次的水平。

图7展示了出生小猫的数量。在 2017 年-2018 年，出生数量的波动量较大，从每 14 天 10 万只到每 14 天 10 只不等；2019 年至 2020 年，每 14 天的出生数量维持在 1 万只左右；但是到了 2022 年，每 14 天的出生数量下降至 100 只左右。

总的来说，CryptoKitties 从 2017 年 11 月 23 号首发，在同年 12 月即达到巅峰，并在 2018 年一整年维持在了一个较高的水平，但是由于其玩法单一，以投机为主，在经历了短暂的高热度后市场活跃度开始逐渐走低，交易量显著减少。

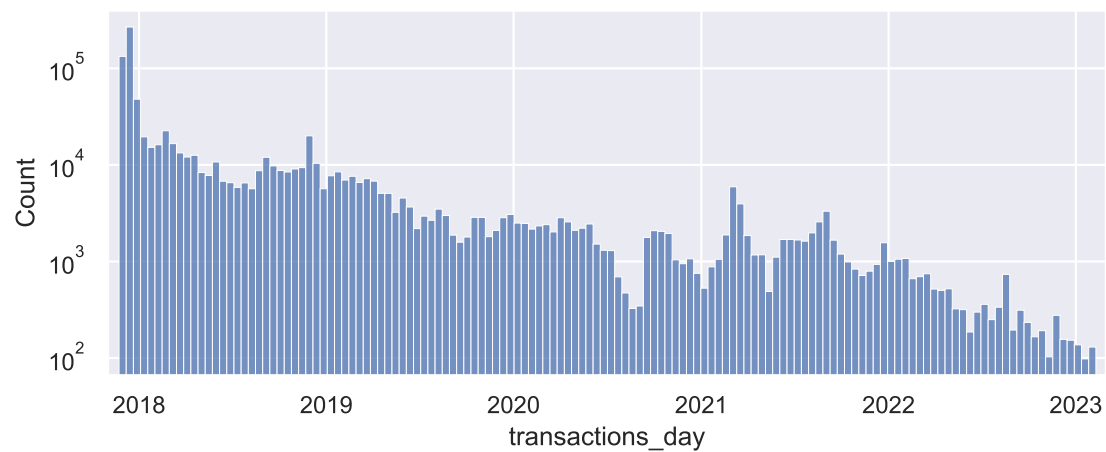


图 4: 销售拍卖成功数量

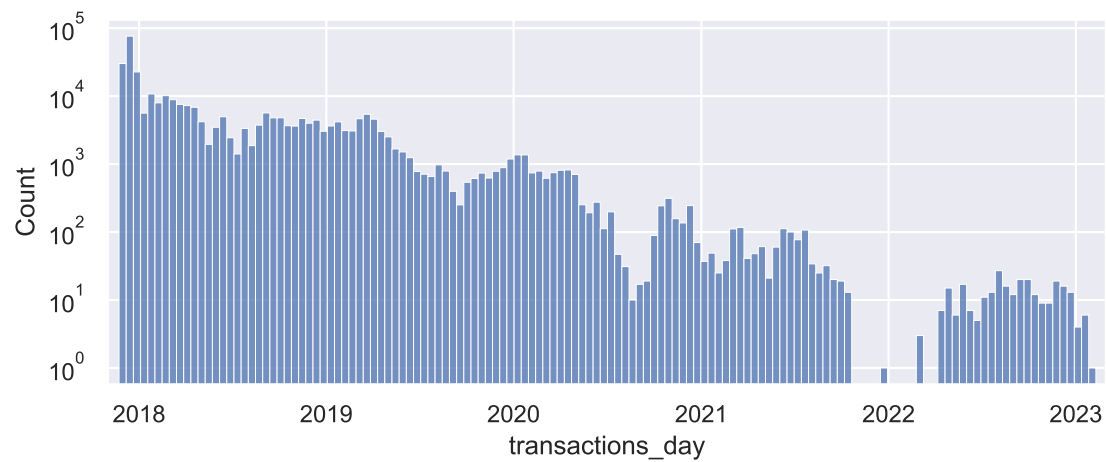


图 5: 创建生育拍卖的频次

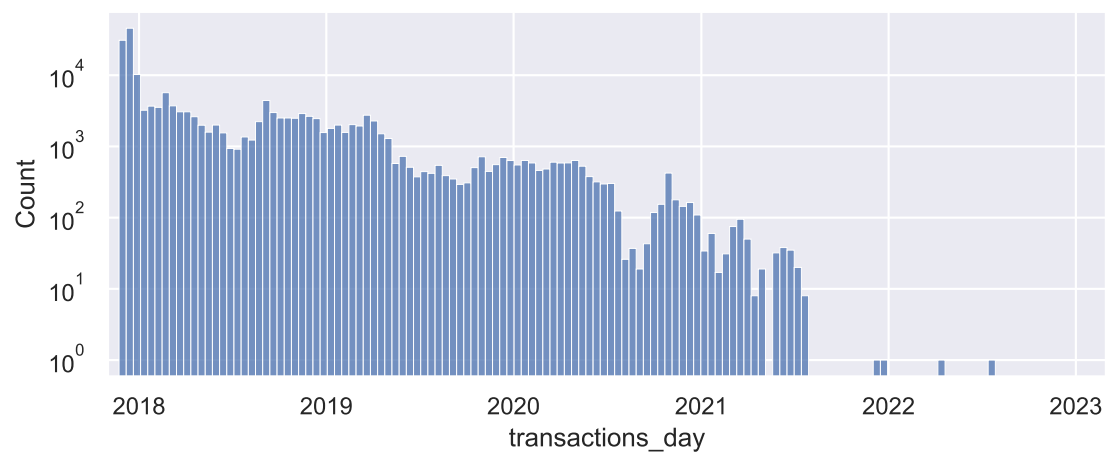


图 6: 生育拍卖成功数量

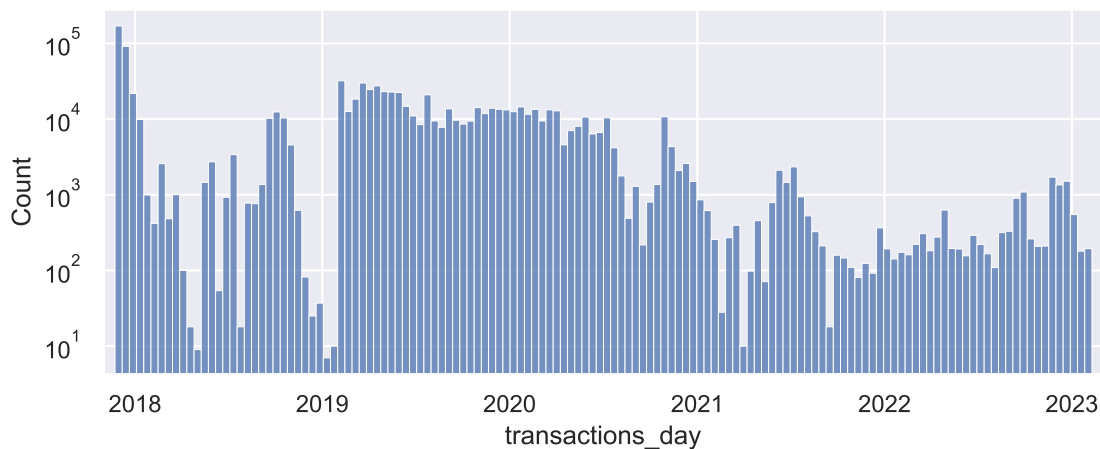


图 7: 出生小猫数量

2. CryptoKitties 的市场价格波动

市场价格波动是指市场价格在一段时间内不断变动的情况。市场价格波动代表了市场供求关系的变化。当市场需求量增加，市场价格会上涨；当市场供给量增加，市场价格则会下降。市场价格波动还反映了市场对于公司、产品、行业的信心度。当市场对公司、产品、行业充满信心，市场价格会上涨；当市场对公司、产品、行业信心不足，市场价格则会下降。

图8-图15展示了从 2017 年 11 月至 2023 年 1 月的市场价格波动的相关数据，其中横轴表示时间，以天为单位；纵轴表示价格，以对数坐标呈现。其中实线表示价格的平均值，半透明的区域表示当天的价格区间。

图8和图9分别展示了以以太坊和以美元结算的拍卖开始价格，其中拍卖开始价格大于 100 亿美元的作为异常值被删掉，没有显示在图中。其中以美元为结算的价格波动图中展示出了较为明显的规律。总体上呈现出先下降、后上升、再下降的趋势。拍卖开始价格在一定程度上表明了用户对于市场的信心，一个较高的拍卖开始价格表明用户更有信心卖出高价。在 2021 年下半年后，拍卖开始价格呈现明显的下降趋势，表明用户对于市场的信心正在逐渐下降，这和市场活跃度的降低相吻合。

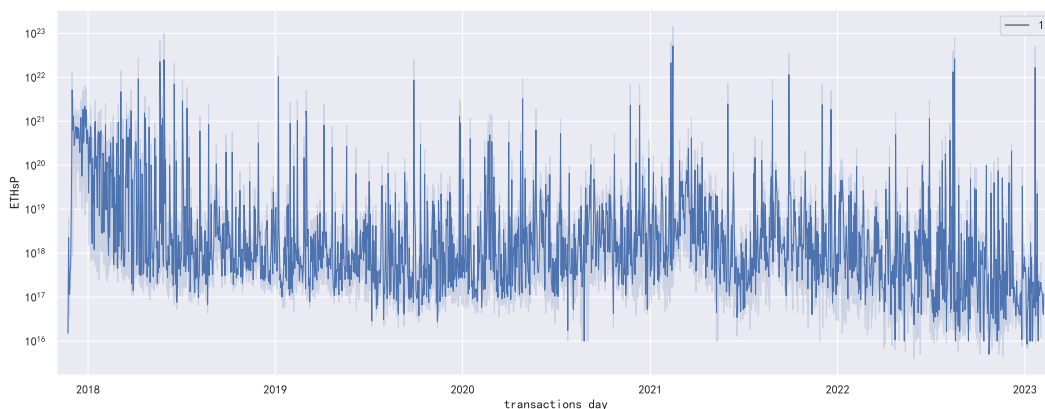


图 8: 拍卖开始价格（Wei）

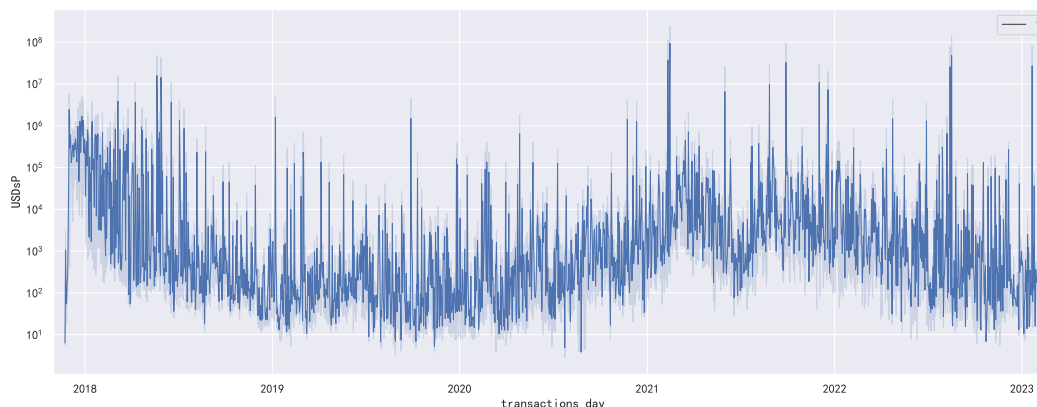


图 9: 拍卖开始价格 (USD)

图11和图12分别展示了以以太坊和以美元结算的拍卖结束价格。其中拍卖结束价格大于 100 亿美元的作为异常值被删掉，没有显示在图中。拍卖结束价格有三种选择：比开始价格低，等于开始价格和比开始价格高，不同选择的比例如图10所示。虽然 CryptoKitties 没有限制开始拍卖价格一定要比结束拍卖价格高，但由于其拍卖模式，有 82.6% 的人选择了荷兰式拍卖（即降价式拍卖），有 12.5% 的人选择了定价拍卖，仅有 4.8% 的人选择了增价拍卖。因此拍卖的结束价格在一定程度上表明了拍卖者对于拍卖的猫的价值判断，而高出结束价格的部分代表某种消费者剩余。如果拍卖价格低于拍卖者的心理预期价值，拍卖者宁可流拍。

和拍卖开始价格类似，拍卖结束价格在以美元结算的情况下体现出了较强的规律性，拍卖结束价格先下降、后上升、再下降。表明 2022 年以来，CryptoKittie 上的猫的价值正在逐渐下降。

但值得注意的是，虽然目前呈现下降趋势，其平均价格仍然比 2019 年的时候高。这主要是受到了供需关系的影响：图13展示了每日创建拍卖数量与竞拍成功数量之差，由于拍卖往往持续数天，所以取值有可能为负数。图中只展示了-500-500 的区间。可以看到一开始由于有大量的猫在市场上流通，所以分布较为分散且远离 0，而到 2022 年以后，每日创建拍卖数量与竞拍成功数量之差在 0 附近波动，这表明需求大于供应，几乎所有的供应都被消化掉了。这种供给短缺造成了价格仍在高位。

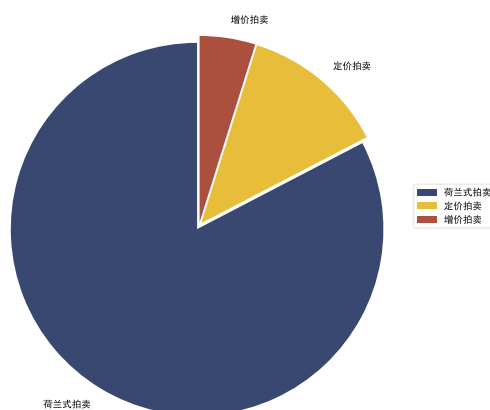


图 10: 不同拍卖的比例

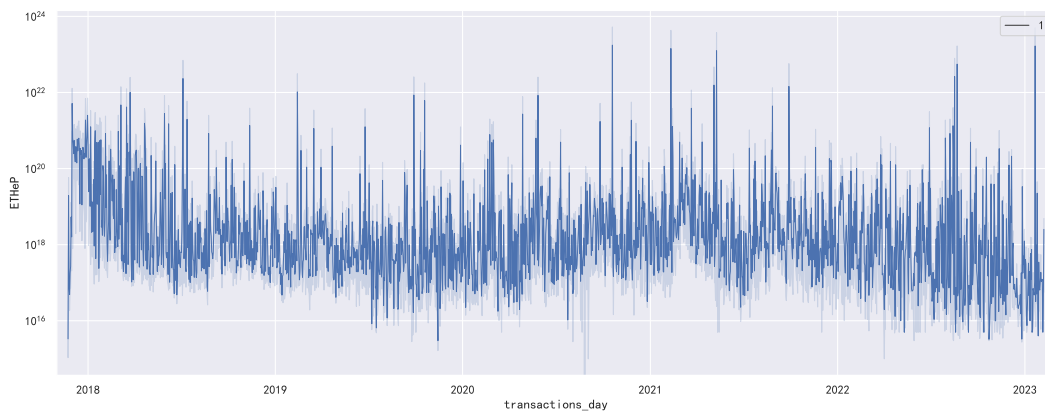


图 11: 拍卖结束价格 (Wei)

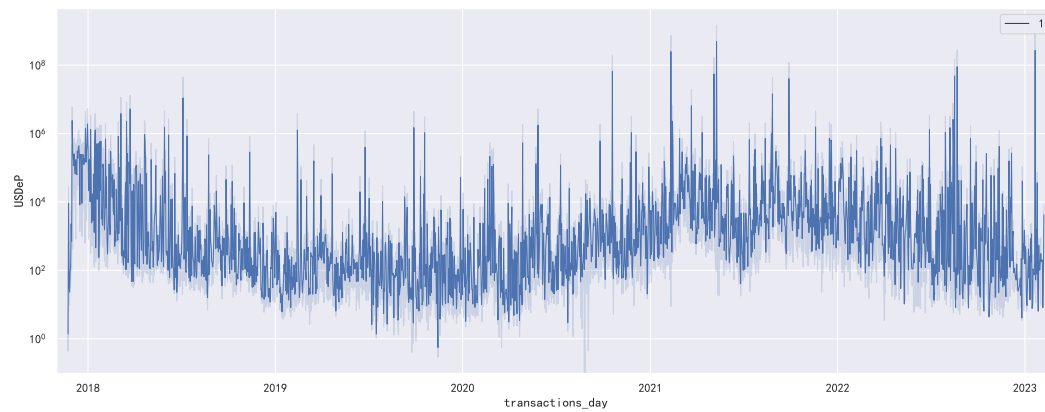


图 12: 拍卖结束价格 (USD)

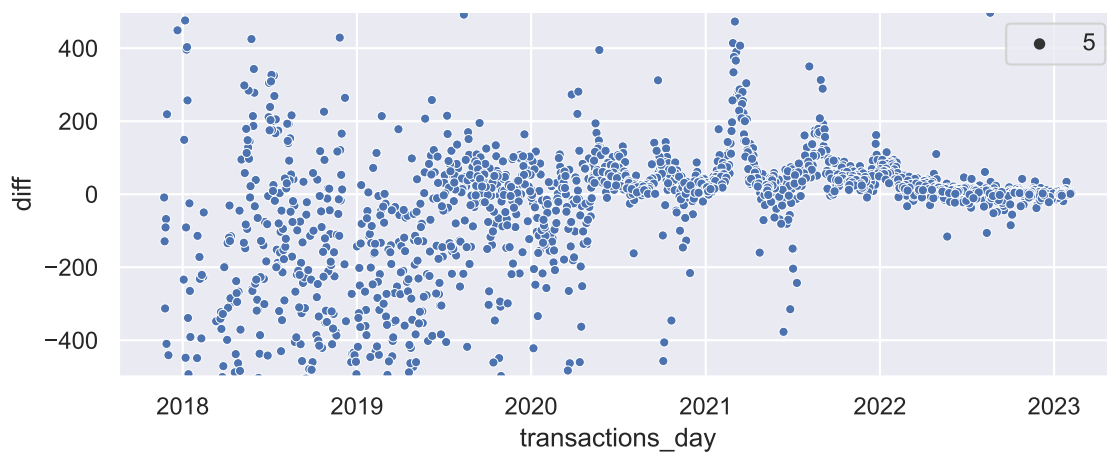


图 13: 每日创建拍卖数量与竞拍成功数量之差

图14和图15分别展示了以以太坊和美元结算的拍卖成交价格。可以看到，2018 年成交价格在 15 美元左右，2019 年跌至 4 美元左右，从 2020 开始到 2021 年下半年，随着以太坊（ETH）价格的突飞猛涨，成交价格也提升到了上百美元左右。2021 年下半年以后，成交价格逐渐回落到 18 年的水平。

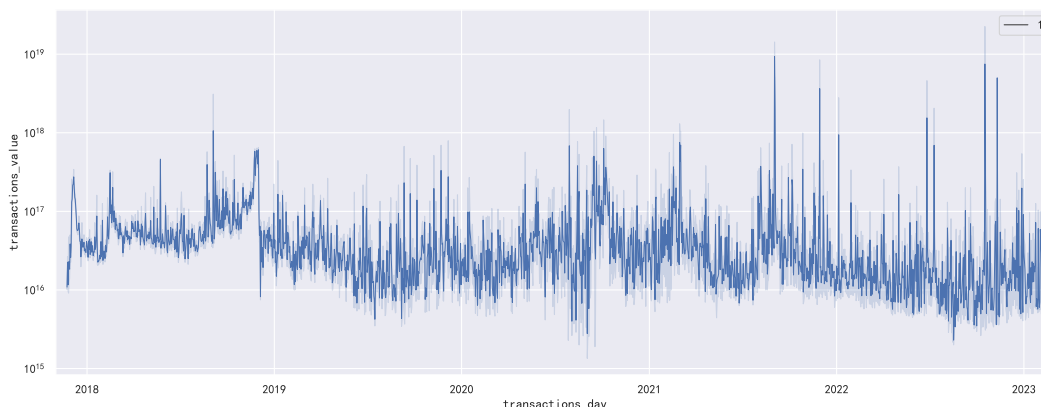


图 14: 成交价格（Wei）

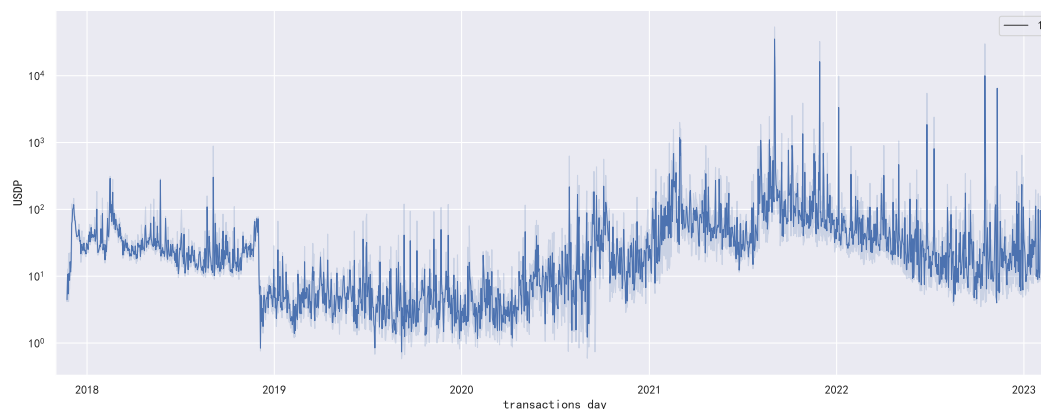


图 15: 成交价格（USD）

总的来说，CryptoKitties 的拍卖价格受多方面影响。如用户对于市场的信心、市场活跃度、供需关系、以太坊价格、猫的个体差异等等。受个体差异影响，成交价格波动较为剧烈，每日成交价的最高价和最低价可能相差 20 倍以上。但是从总体趋势而言，呈现先下降、后上升、再下降的趋势。并且虽然当前市场活跃度较低，价格也在走低，但是不论是拍卖开始价格、拍卖结束价格还是成交价格仍高于 2019-2020 年的价格。

3. CryptoKitties 的用户分析

除了 CryptoKitties 的市场活跃度和市场价格变化外，用户交易频次的信息也十分重要。图16-图18展示了用户调用 Core 和 SaleAuction 两个合约中的函数的频次。由于调用函数在 100 以上的用户数量较少，所以仅展示了调用函数次数在 100 以内的数据。从图16中可以看到，有将近 17,000 名

用户仅仅调用了一次，用户数量随着调用次数的增加而迅速减少。这表明大部分人将 CryptoKitties 当做一个非同质化代币（NFT）产品或者一款游戏，仅有小部分人以此牟利。图17展示了核密度估计的不同交易频次的用户数量的分布情况。

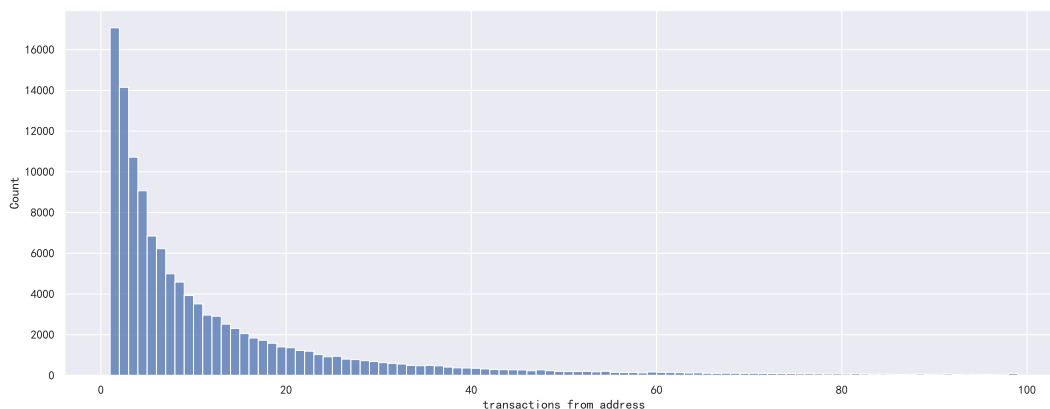


图 16: 用户调用函数的次数

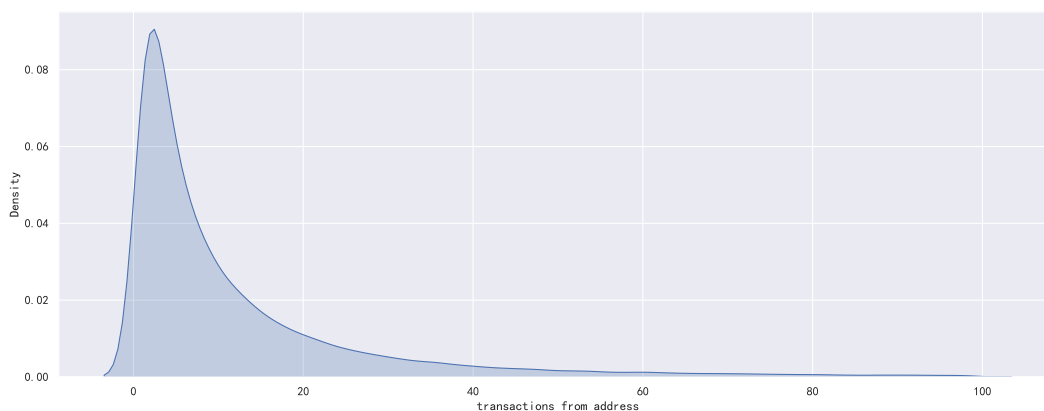


图 17: 用户调用函数的次数的核密度估计

图18展示了不同调用频次区间的用户数量所占比例。区间以指数划分 (从 e^1 到 e^4)。其中，1-3 次占比 32.9%，4-8 次占比 24.9%，9-20 次占比 22.1%，21-55 次占比 13.3%，55 次以上占比 6.8%。

不同调用次数的用户群体中，其调用函数的占比也有所区别。图19显示了不同调用次数下，调用频次大于总调用数量 5% 的函数的占比。可以看到，函数 0x454a2ab3 (bid) 的调用比例随着调用次数的增加而逐渐减少，函数 0xf7d8c883 (breedWithAuto) 的调用比例随着调用次数的增加而增加。0x3d7d3f5a (createSaleAuction) 的调用比例随着调用次数的增加而增加，但是在大于 55 次的部分有所减少。0xed60ade6 (bidOnSiringAuction) 在 9-55 次的部分占比较多，其原因在于良好的种猫往往在重度玩家手中，且价格昂贵，轻度玩家如果想获得较好的基因则需要购买生育权。最后，0x88c2a0bf (Birth) 的调用次数仅在调用频次大于 55 的群体中占比高于 5%，且以 18.8% 的占比成为调用频次第二高的函数。

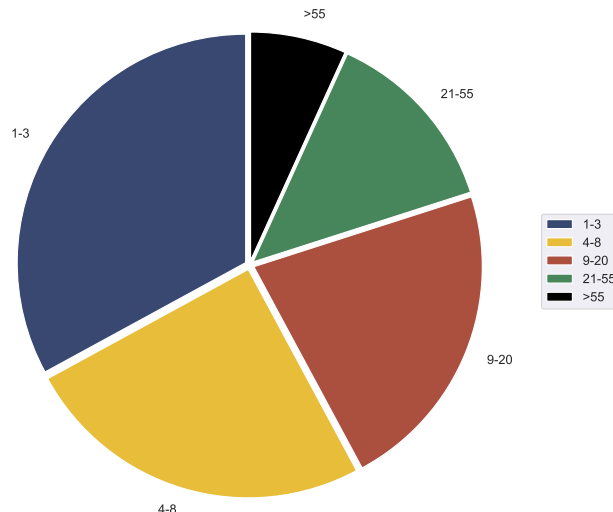


图 18: 不同调用次数区间的比例

4. CryptoKitties 的用户行为分析

用户行为分析指的是通过对用户在使用产品和服务时的行为进行研究和分析，来了解用户的需求和行为。

在本节中，我们选取了地址为 `0x85a2_9ad3`¹ 的用户进行分析，该用户在 2017 年至 2023 年间总共调用了 1497 次 Core 和 SaleAuction 上的函数。在本节，我们尝试在回答五个问题：

- (1) 该用户购买了哪些猫？
- (2) 该用户购买了哪些猫的交配权？
- (3) 该用户创建了哪些拍卖？
- (4) 该用户创建的哪些拍卖成功了？
- (5) 该用户自己的猫之间是如何繁衍的？

图20和图21展示了该用户的行为模式图，其中绿色节点表示以太坊账户，橙色节点表示 CryptoKitties 中的小猫。深蓝色箭头表示用户对猫创建一个拍卖，箭头由用户指向猫；蓝色箭头表示猫被出售给某个用户，箭头由猫指向用户，不是所有的创建拍卖都会拍卖成功。橘红色的粗箭头表示用户购入某个猫，箭头由猫指向用户；红色箭头表示购买生育权用来给用户自己的猫配种，箭头由种猫指向用户的猫；黄色箭头表示用户自己的猫之间发生交配，箭头由“父亲”指向“母亲”。

整张图采用 spring layout 的布局算法。spring layout 是一种基于力导向图（force-directed graph layout algorithm）布局的算法，通过将节点视为物理对象，并计算它们之间的斥力和吸引力来生成图形布局。其中距离越近的点表明其关系越紧密，距离越远的点表明其关系越稀疏。

¹完整地址：`0x85a243ceb8539884f0aa935408256c7f37c79ad3`

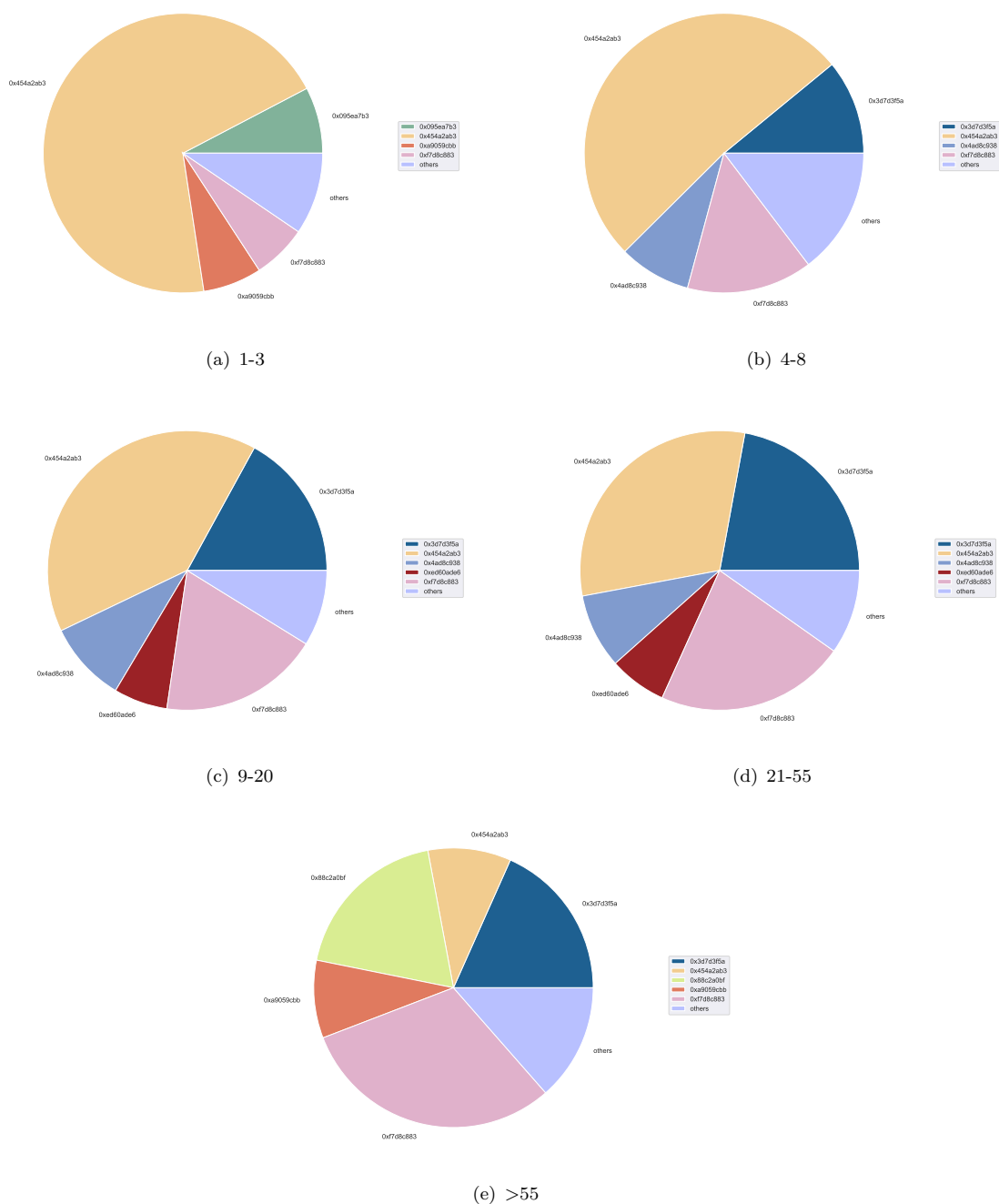


图 19: 不同调用次数区间的用户调用函数的比例

图20中中心较为密集的点关系比较“亲密”，往往与其它节点有着多种关系，如购买、生育、发起拍卖、拍卖成功等等。而外圈与以红色箭头指向中心的节点大部分只有购买和拍卖量两种关系。最外圈的节点则不存在购买关系，只存在拍卖关系和拍卖成功关系。

图21中可以看到中心部分的详细关系。用户买入小猫，并让自己的小猫相互之间进行大量繁殖，同时也购买一定的生育权以获得更好的基因。在繁殖之后，生成的小猫被用户拿去售卖，480次拍卖中拍卖成功了434次，有较为可观的成功率与收入。

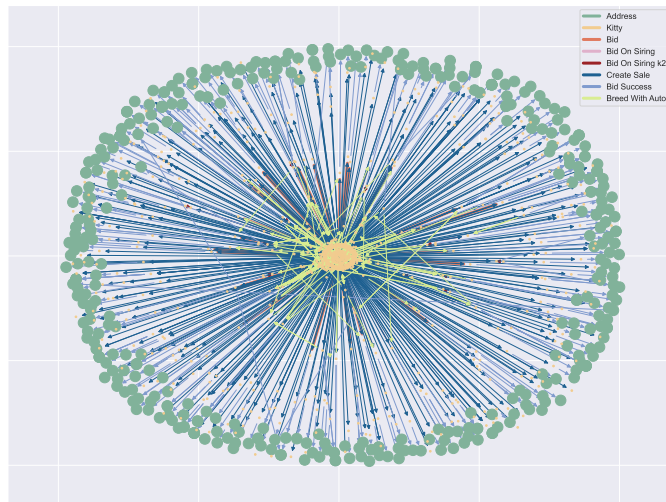


图 20: 用户行为模式分析

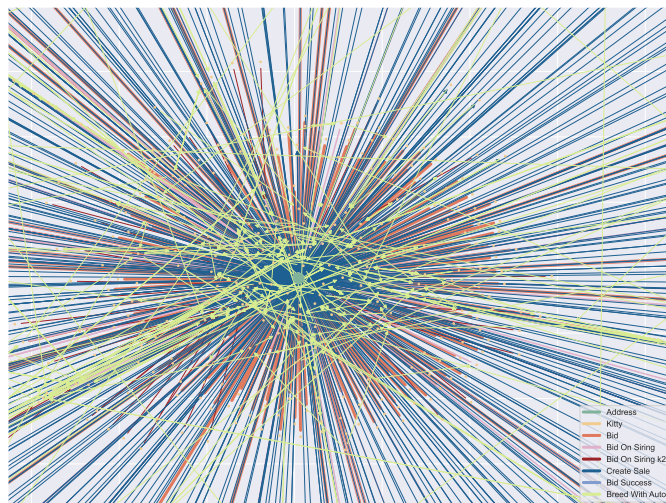


图 21: 用户行为模式分析

总的来说，交易频次高的用户通过购买初代猫或其它优质的猫获得原始资本，并通过繁殖产生新的小猫。一只小猫的价格一般在 0.05ETH 左右，而产生小猫需要使用 10 万左右的 gas，约合 0.002ETH。期间的差价即为用户的盈利。在以太坊发布完所有的初代猫以后，手握大量猫的用户就成为了新的“发行商”。如果 CryptoKitties 的市场不断扩大，则会有大量用户涌入并购买一只或多只小猫，卖出去的猫越多，用户的盈利就越高；而如果市场发展不佳，则有可能造成入不敷出的局面。

5. CryptoKitties 中的猫的价格预测

CryptoKitties 市场中，猫的交易价格是一个非常重要的因素，它受到多种因素的影响，如稀有度、基因组合、出生日期等，在本节中，我们选取了 2018 年一整年的交易数据，共计 292829 条，通过机器学习 and 数据挖掘技术，对猫的交易价格进行预测。

该预测结果可以帮助 CryptoKitties 的玩家和投资者更好地了解市场行情，以制定合适的拍卖开始价格和结束价格。

5.1 数据预处理

生成猫的信息


猫的出生有两种形式：一种是 CryptoKitties 官方所创建的初代猫，COO 每隔 15 分钟调用 createGen0Auction 函数产生一个 0 代的猫，并进行拍卖。另一种则是用户通过交配、怀孕，最后通过 giveBirth 函数生成一只新的小猫。这两种情况都会产生一个名为 Birth 的 Event Log，其中记录了新出生的猫的所有者、Id、母亲的 Id、父亲的 Id 和这只猫的基因。从这些数据中，我们可以获取到每只猫的基因、猫的代数以及每只猫每次繁衍的时间。

其中，猫的代数为其父亲的代数和母亲的代数的最大值加一，即

$$\text{kitty_generation} = \max(\text{matron_generation}, \text{sire_generation}) + 1$$

猫的代数影响了猫生育后的休息时间，如表2所示。此外当这只猫繁衍后，它的特征会发生变化。每次繁衍等同于增加了一个代。如果一只 3 代猫繁衍了 3 次，它的修养特征会变成第 6 代的 Snappy。这样就限制了一只猫的繁衍速度，增加了低代的稀有价值，也限制了稀有基因的遗传效率。

猫的基因由一个 64 位的 16 进制数表示，将其转换成 256 的 2 进制数后，去掉前 16 位的 0，剩余 240 位，每 5 位为一组，对应了 48 个基因位，用 Kai 编码表示。48 个基因位中每 4 个位一组，形成 12 块，如图22所示。在每一个基因块中，第四位表示了当前猫的显性性状，其它倒数第二位、第三位、第四位分别表示了第一、第二、第三隐性性状。



Kitty #9181

Owner: 0x0ccffF882e82654A73F15Af88A02276776d11c51
Born: November 27, 2017 at 2:06 PM UTC
Generation: 12
Block number: 0x46ae18
Block hash: 0x1ac953133f320337679e4b9b6a5c82d75705742e9907eb808cda6c918404ff4e
Official profile: [CryptoKitty #9181](#)
Cattributes: happygokitty, munchkin, luckystripe, sizzurp, granitegrey, simple, cloudwhite, lemonade

Genes

Use the explorer below to browse this kitty's genes in different formats. The most useful format is named after @kaigani, the CryptoKitty scientist who mapped the genome.

[kai](#) [hex](#) [binary](#)

a99m 7676 222f 9g9f 4644 7575 eFee bb7h 6866 8646 faga fccd

图 22: 基因编码

表 2: 不同代的猫的初始生育休息时间

代	特征	修养时间
0 • 1	Fast	1 分钟
2 • 3	Swift	2 分钟
4 • 5	Swift	5 分钟
6 • 7	Snappy	10 分钟
8 • 9	Snappy	30 分钟
10 • 11	Brisk	1 小时
12 • 13	Brisk	2 小时
14 • 15	Plodding	4 小时
16 • 17	Plodding	8 小时
18 • 19	Slow	16 小时
20 • 21	Slow	24 小时
22 • 23	Sluggish	2 天
24 • 45	Sluggish	4 天
26+	Catatonic	1 周

显然，猫的基因、代和交易前的繁衍次数都会影响其交易价格。在这一部分中，我们选取了 2018 年以前 Log 数据库中合约地址为 Core 且 topic 为 Birth 的所有数据（2018 年的交易数据可能包含 2018 年以前生成的猫），通过解析 log data 得到 kittyId、matronId、sireId、gene 四个字段的信息，并获取 log 所在的区块的编号 block_number 作为出生时间。由于每只猫只会出生一次，因此用猫的 Id 作为索引²。随后添加 pregnant 字段，遍历整张表，在每只小猫的母亲的可孕字段中添加小猫的出生区块号。最后添加 generation 字段，将初代猫初始化为第 0 代，通过递归的方式计算每只猫的代数。

生成市场信息

除了猫本身，市场的活跃度和供需关系显然也会影响到猫的市场价格。因此和研究主题一中类似的，统计每日创建销售拍卖的次数、创建生育拍卖的次数、销售拍卖成功的数量和生育拍卖成功的数量和 birth 和 breedWithAuto 的调用次数。由于供需关系（bid-createSaleAuction）是 bid 和 createSaleAuction 的线性组合，因此舍去。

生成销售数据

以 Sale Auction 中调用函数为 bid 的数据为基础，从 input 中解析出竞拍的小猫的 Id，通过 Id 连接包含所有猫的数据的表 kitty_table。再通过日期连接市场活跃度的表 market，和以太币价格的表 price，从而得到 sale_data。

遍历 sale_data，对于每个竞拍成功的数据，从 Core 中通过 kittyId 找到 createSaleAuction 的候选者，再通过日期的关系找到 bid 对应的 createSaleAuction。并获取其初始价格、结束价格和持续时间。

最后通过猫的可孕的区块编号列表和 bid 时的区块编号计算在拍卖之前，该猫一共生育过多少次。并将 48 位 kai 编码的基因序列转换为 48 个类别变量，并保存。

变量筛选与变量处理

- (1) 计算以美元结算的交易价格。
- (2) 删除所有数值变量和类别变量以外的变量，如 transactions_hash、transactions_input 等等。

²加快索引速度

- (3) 删除与预测竞拍价格无关的变量,如 transactions_block_number、transaction_index、start_price、end_price、transactions_value 等等。
- (4) 将类别变量转换为独热编码。首先将 48 个基因位转换为独热编码, 另外由于代 (generation) 是离散且非线性变化的, 因此代也作为类别变量被转换为独热编码。
- (5) 将左偏的数据进行对数化处理, 使数据分布更加接近正态分布, 更符合线性回归的假设。所有数值数据的分布情况如图23所示。可以看到, 竞拍价格 (bid price)、持续时间 (duration)、生育次数 (pregnant count)、油费 (gas_price) 有较为严重的左偏现象。因此对这些变量进行对数化处理。

最终得到了共计 1470 个变量。

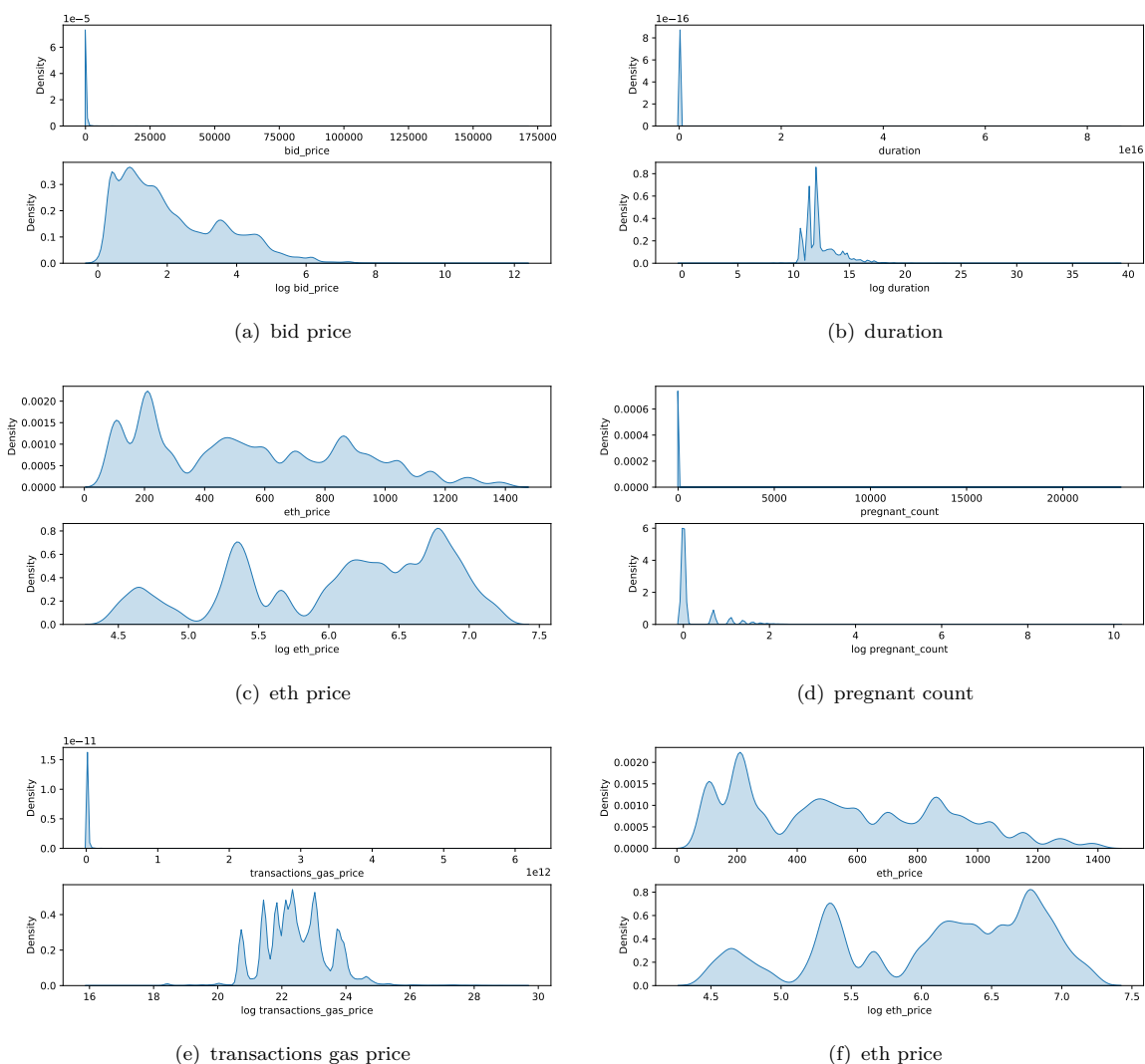


图 23: 数值变量分布

5.2 运行结果

经过线性回归模型的拟合后, MSE 为 0.664, R^2 为 0.703, 可解释能力较强。

其中，价格随着交易时间、持续时间、创建销售拍卖的数量、以太坊价格的增加而增加，随创建生育拍卖的数量和汽油费的增加而下降。

基因部分的系数如图24所示。其中红色代表系数为正，蓝色代表系数为负。所有系数被限制在-1 到 1 之间。

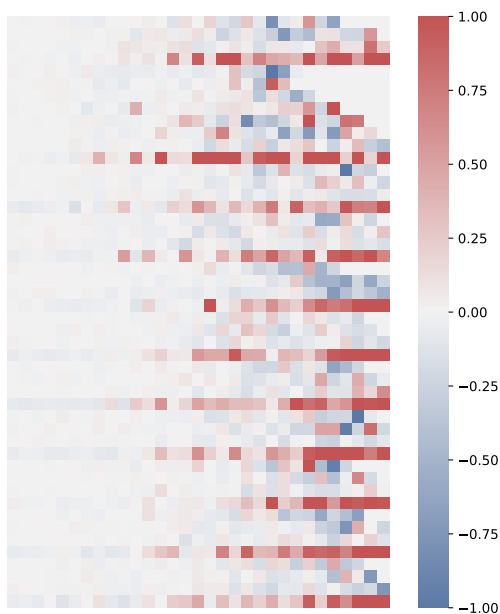


图 24: 基因对于拍卖成交价格的影响

对于 generation，其系数变化如图25所示。可以看到从 0 代猫到 3 代猫之间价格迅速下降，从 5 代猫到 15 代猫之间猫的代数对于价格没有太大影响，15 代以后又有小幅下降。

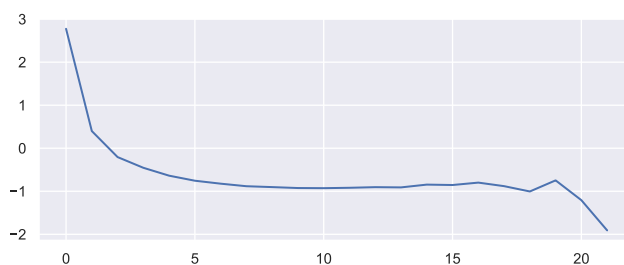


图 25: 代数对于拍卖成交价格的影响

五、 总结与展望

CryptoKitties 是以太坊 DApp 的先驱者，但是近年来，CryptoKitties 的市场活跃度在逐渐下降，市场价格先下降、后上升、再下降，市场目前需求量与供给量持平或大于供给量。CryptoKitties 的大

部分用户以收藏 NFT 为主要目的，少部分人以牟利为主要目的。用户在 CryptoKitties 的盈利模式为购入种猫、让自己的猫繁殖或购买生育权繁殖、将繁殖后出生的小猫或其它手中的猫再卖出去，以赢得利润。CryptoKitties 市场中的交易价格受到多方面的因素影响。在猫本身方面，包括猫的基因、猫的代数、猫的生育次数等等。在市场方面，包括创建拍卖的次数、拍卖成功的次数、以太币价格等等。此外还受到时间等因素的影响。通过线性模型拟合 2018 年的数据，其 R^2 为 0.703，可解释性较高。

线性模型需要基于一定的假设。若拍卖价格等变量没有进行对数化处理，则 R^2 仅有 0.1 左右，即使进行对数化处理，也仅仅假设该变量的对数满足线性化假设。然而在实际过程中，解释变量与被解释变量之间的关系可能远比指数更复杂。因此在未来，可以尝试加入交叉项或其它函数形式来丰富模型的假设，或者使用深度学习的方法，让模型学习更加复杂的特征。

参考文献

- [1] MECO CLUB. *WEB 3.0 时代下的音乐产业洞察报告*. 2022.